How to Test for Heteroscedasticity with the Breusch-Pagan Test in SPSS

Authored by

mohammed looti

January 7, 2026

RECOMMENDED CITATION

mohammed looti (2026). *How to Test for Heteroscedasticity with the Breusch-Pagan Test in SPSS*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=124812

The Breusch-Pagan test is an essential diagnostic tool in statistical analysis, particularly within linear regression. Its primary function is to check for the presence of heteroscedasticity—a condition where the variance of the residuals (errors) is not constant across all levels of the independent variables. Homoscedasticity (constant variance) is one of the crucial assumptions for validating inferences drawn from a standard Ordinary Least Squares (OLS) regression model.

If heteroscedasticity is present, the parameter estimates remain unbiased, but the standard errors become biased, leading to inaccurate confidence intervals and potentially incorrect hypothesis testing results. Therefore, performing the Breusch-Pagan test is a critical step. This guide provides an expert, step-by-step walkthrough detailing how to execute and interpret this important diagnostic test using the SPSS statistical software package.

The Breusch-Pagan test is designed to formally assess whether the variance of the errors in a regression is dependent on the values of the independent variables. This procedure involves running an auxiliary regression using the squared residuals obtained from the primary model.

The following comprehensive, step-by-step example illustrates the correct process for executing the Breusch-Pagan test in SPSS, ensuring the validity of your regression findings.

Step 1: Setting up the Data and Initial Model Formulation

To demonstrate the procedure, we will use a hypothetical scenario involving a multiple linear regression analysis. Suppose our objective is to fit a linear regression model that utilizes two predictor variables—the number of hours spent studying and the number of prep exams taken—to predict the final exam score of students.

The theoretical structure of this model is defined as:

Exam Score = β0 + β1(hours) +β2(prep exams) + ε

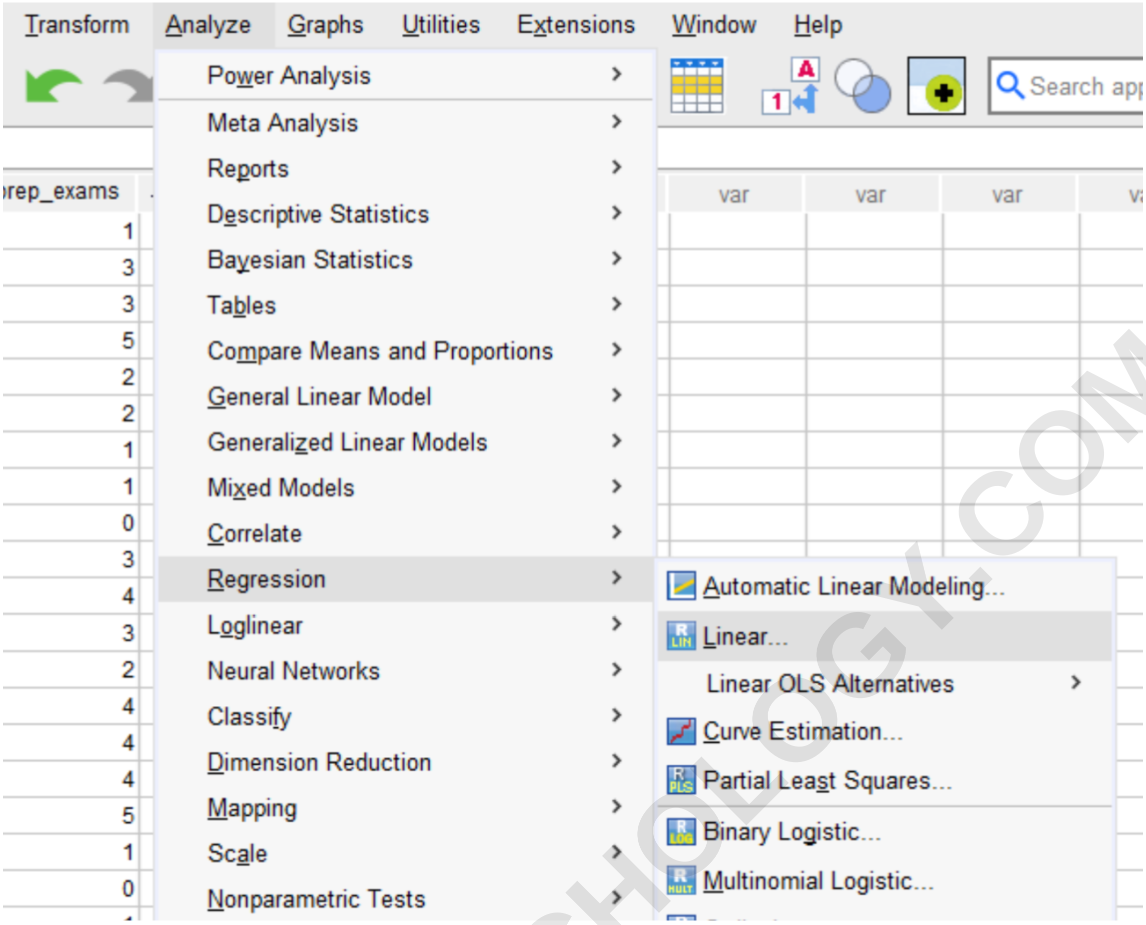
Our initial task is to input the relevant dataset into SPSS. This dataset contains the scores and predictor information for 20 unique students, which will serve as the foundation for our analysis:

	hours	prep_exams	score	var	var
1	1	1	76		
2	2	3	78		
3	2	3	85		
4	4	5	88		
5	2	2	72		
6	1	2	69		
7	5	1	94		
8	4	1	94		
9	2	0	88		
10	4	3	92		
11	4	4	90		
12	3	3	75		
13	6	2	90		
14	5	4	90		
15	3	4	82		
16	4	4	85		
17	6	5	90		
18	2	1	83		
19	1	0	62		
20	2	1	76		
21					
22					
23					

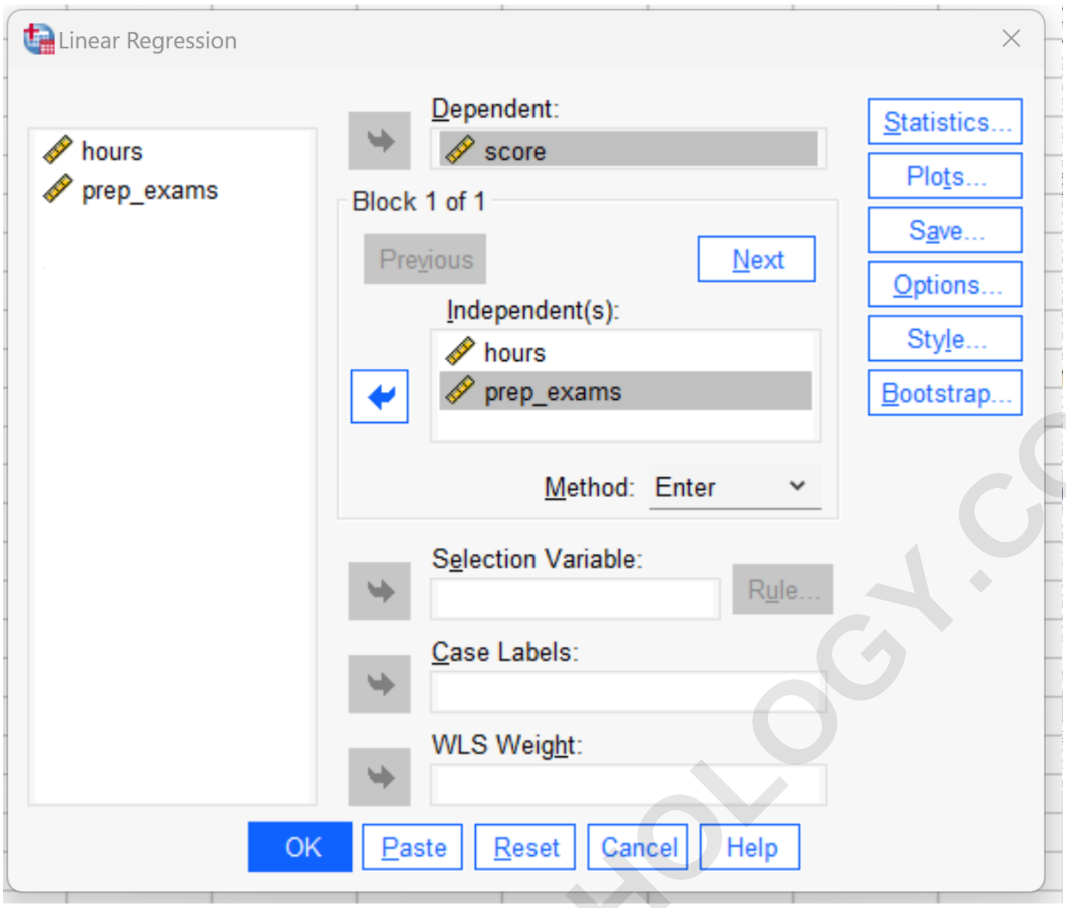
Step 2: Fitting the Primary Linear Regression Model

Once the data is correctly structured in SPSS, the next critical step is to fit the primary multiple linear regression model. Furthermore, we must instruct SPSS to save the essential components required for the Breusch-Pagan calculation: the unstandardized predicted values and the unstandardized residuals.

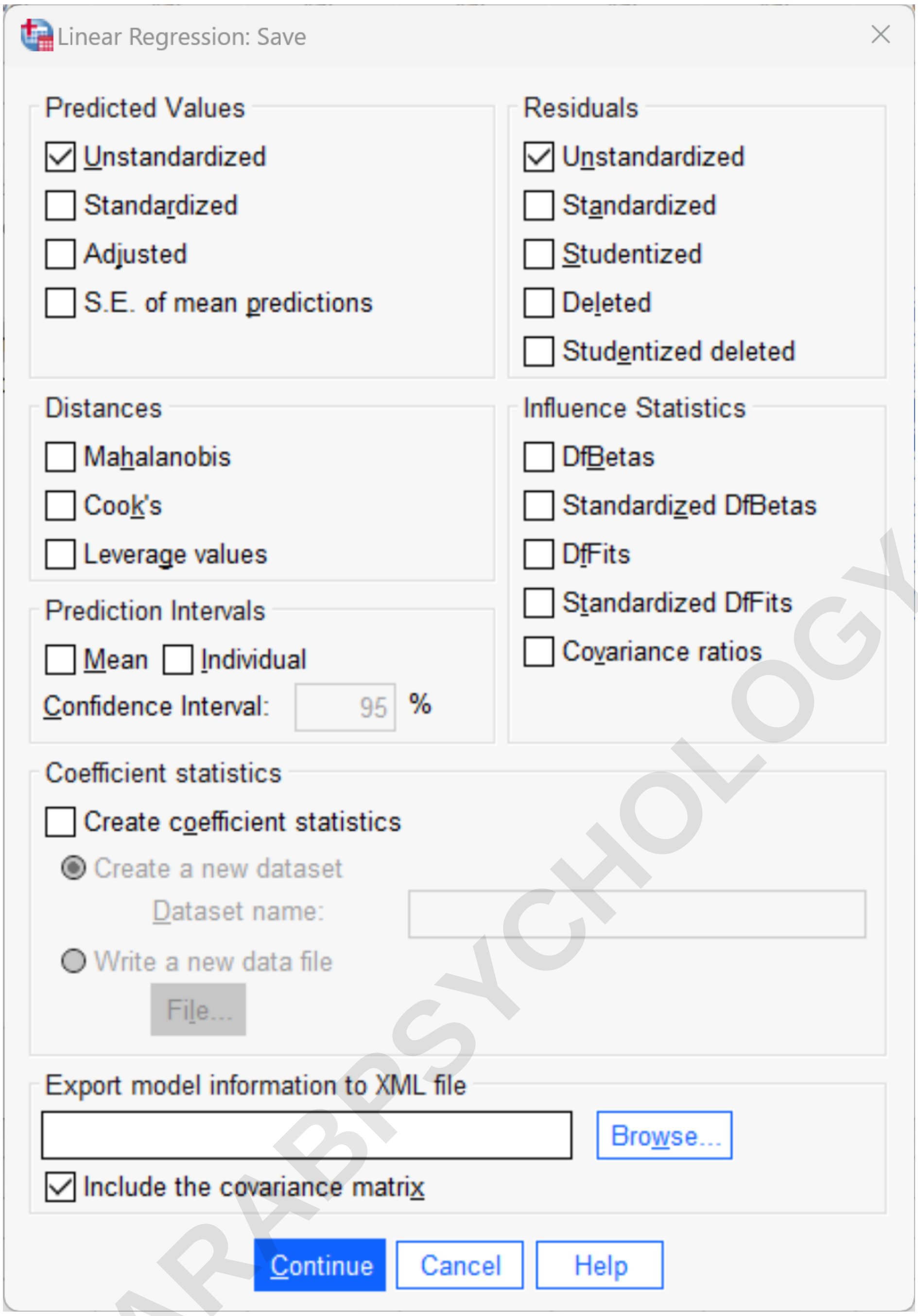
To initiate the regression, navigate through the menu bar: Click the **Analyze** tab, hover over **Regression**, and then click **Linear**. This action opens the Linear Regression dialogue box:



In the dialogue box that appears, specify the model variables: Drag the **score** variable into the **Dependent** panel (as it is the variable we are predicting), and then drag both **hours** and **prep_exams** into the **Independent(s)** panel (as they are the predictors):



After specifying the variables, click the **Save** button located within the Linear Regression dialogue box. This opens the submenu for saving regression output variables to the dataset. Under the **Predicted Values** section, check the box next to **Unstandardized**. Similarly, under the **Residuals** section, check the box next to **Unstandardized**. These two new variables, often labeled PRE_1 and RES_1, are crucial for the subsequent steps of the Breusch-Pagan test:



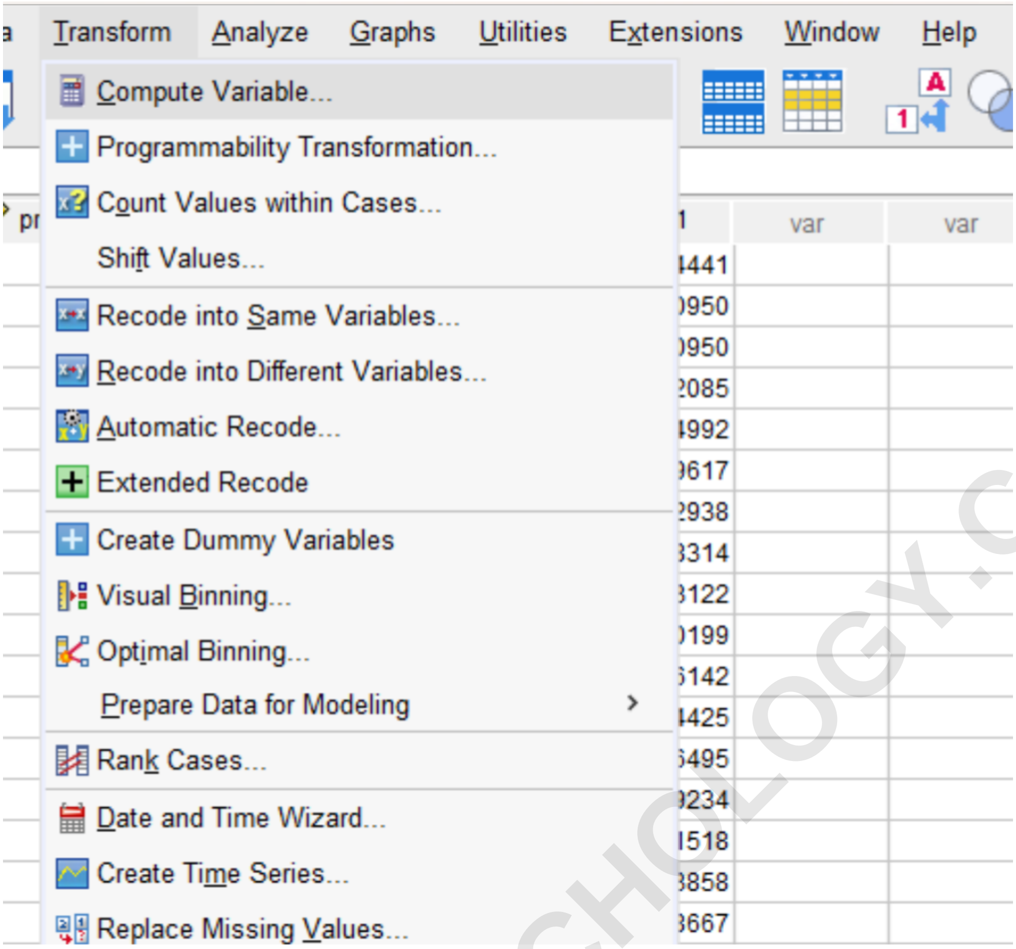
After checking the necessary boxes, click **Continue** to close the Save submenu, and then click **OK** in the main Linear Regression dialogue box to run the analysis. SPSS will generate the standard regression output and add two new columns to your data sheet (predicted values and residuals).

	hours	prep_exams	score	PRE_1	RES_1
1	1	1	76	73.75559	2.24441
2	2	3	78	77.29050	.70950
3	2	3	85	77.29050	7.70950
4	4	5	88	85.47915	2.52085
5	2	2	72	77.84992	-5.84992
6	1	2	69	73.19617	-4.19617
7	5	1	94	92.37062	1.62938
8	4	1	94	87.71686	6.28314
9	2	0	88	78.96878	9.03122
10	4	3	92	86.59801	5.40199
11	4	4	90	86.03858	3.96142
12	3	3	75	81.94425	-6.94425
13	6	2	90	96.46495	-6.46495
14	5	4	90	90.69234	-.69234
15	3	4	82	81.38482	.61518
16	4	4	85	86.03858	-1.03858
17	6	5	90	94.78667	-4.78667
18	2	1	83	78.40935	4.59065
19	1	0	62	74.31502	-12.31502
20	2	1	76	78.40935	-2.40935
21					
22					
23					

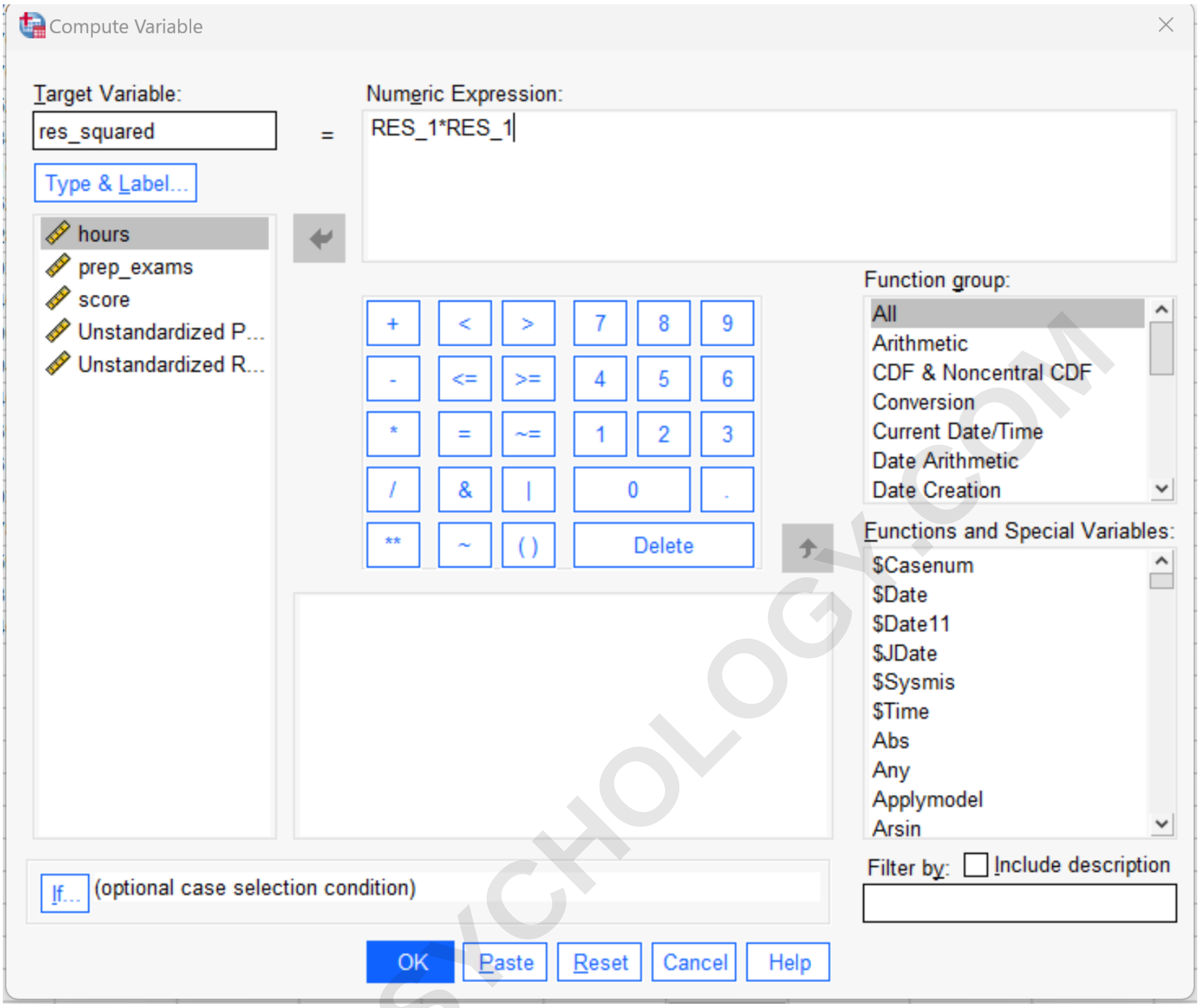
Step 3: Calculating the Squared Residuals for the Test

The core of the Breusch-Pagan test is to determine if the variability of the errors is systematically related to the independent variables. To measure this variability, we must calculate the squared values of the unstandardized residuals (RES_1) saved in the previous step. This calculated variable will serve as the dependent variable in our auxiliary regression.

To create this new variable, click the **Transform** tab in the menu bar and then click **Compute Variable**:



In the Compute Variable dialogue box, assign a suitable name to the new variable, such as **res_squared**, in the **Target Variable** field. For the calculation, type the formula **RES_1*RES_1** in the **Numeric Expression** box. This expression instructs SPSS to multiply the unstandardized residual variable by itself, effectively squaring each residual value:



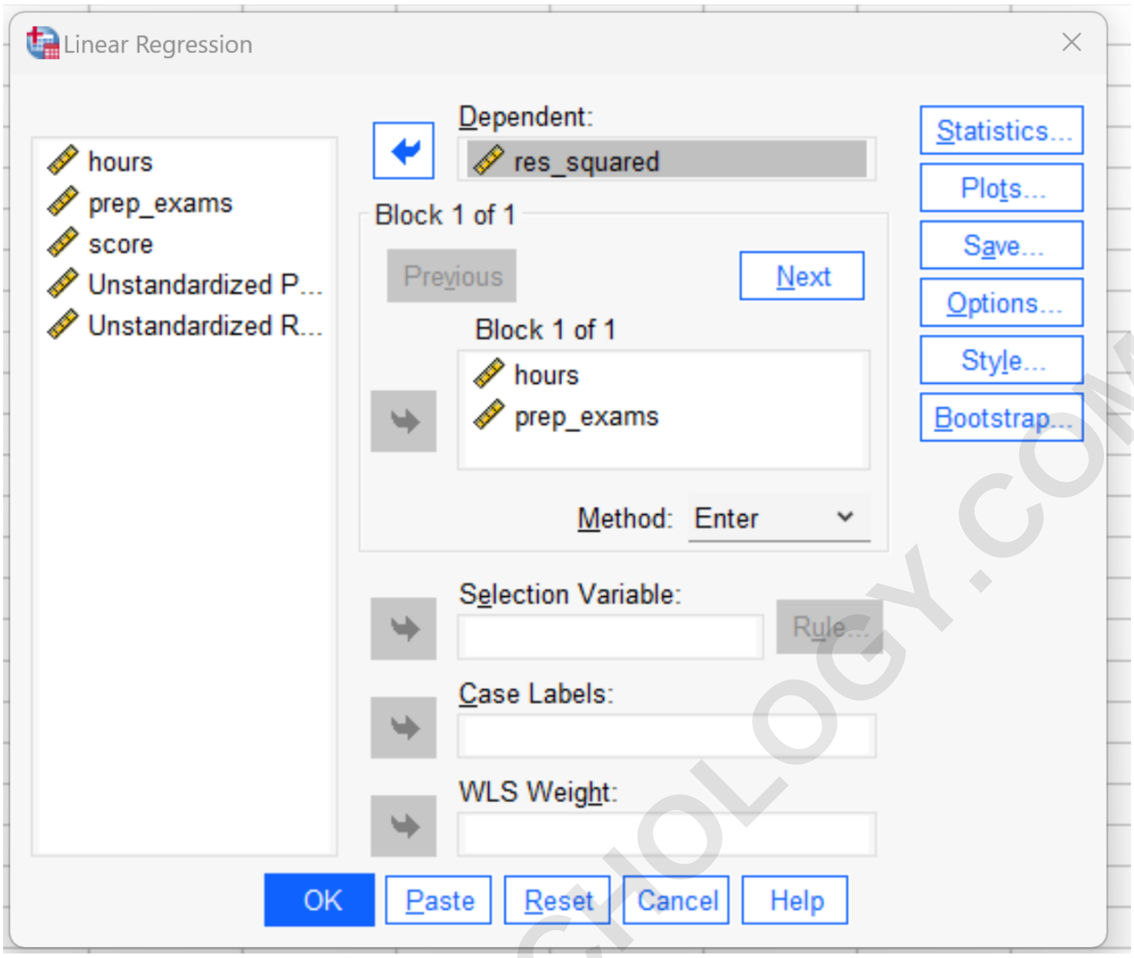
Click **OK** to execute the computation. SPSS will generate a new column in your dataset named **res_squared**, which contains the necessary values for the final test stage. This variable represents the magnitude of the errors for each observation:

	hours	prep_exams	score	PRE_1	RES_1	res_squared
1	1	1	76	73.75559	2.24441	5.04
2	2	3	78	77.29050	.70950	.50
3	2	3	85	77.29050	7.70950	59.44
4	4	5	88	85.47915	2.52085	6.35
5	2	2	72	77.84992	-5.84992	34.22
6	1	2	69	73.19617	-4.19617	17.61
7	5	1	94	92.37062	1.62938	2.65
8	4	1	94	87.71686	6.28314	39.48
9	2	0	88	78.96878	9.03122	81.56
10	4	3	92	86.59801	5.40199	29.18
11	4	4	90	86.03858	3.96142	15.69
12	3	3	75	81.94425	-6.94425	48.22
13	6	2	90	96.46495	-6.46495	41.80
14	5	4	90	90.69234	-.69234	.48
15	3	4	82	81.38482	.61518	.38
16	4	4	85	86.03858	-1.03858	1.08
17	6	5	90	94.78667	-4.78667	22.91
18	2	1	83	78.40935	4.59065	21.07
19	1	0	62	74.31502	-12.31502	151.66
20	2	1	76	78.40935	-2.40935	5.80
21						
22						

Step 4: Executing the Auxiliary Regression and Interpreting Results

The final step of the Breusch-Pagan test involves running a second, or auxiliary, regression model. In this model, the squared residuals (the **res_squared** variable) are used as the dependent variable, while the original predictor variables (hours and prep_exams) are used as the independent variables.

Reopen the Linear Regression dialogue box: Click **Analyze**, then **Regression**, then **Linear**. Now, reset the variables. Drag **res_squared** into the **Dependent** panel. Retain **hours** and **prep_exams** in the **Independent(s)** box:



Click **OK**. The resulting output provides the statistics needed for the Breusch-Pagan calculation, particularly focusing on the ANOVA table from this auxiliary regression:

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	prep_exams, hours[b]	.	Enter

a. Dependent Variable: res_squared

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.502[a]	.252	.164	33.39438

a. Predictors: (Constant), prep_exams, hours

b. Dependent Variable: res_squared

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6395.629	2	3197.814	2.868	.085[b]
	Residual	18958.141	17	1115.185		
	Total	25353.770	19			

a. Dependent Variable: res_squared

b. Predictors: (Constant), prep_exams, hours

The crucial value for the Breusch-Pagan test is the p-value found under the **Sig.** column of the **ANOVA** table for this auxiliary regression. This p-value relates directly to the F-statistic of the auxiliary model. In this specific example, we observe that the p-value is **0.085**.

Step 5: Statistical Interpretation of the Breusch-Pagan Test

The interpretation of the Breusch-Pagan test hinges on the comparison between the calculated p-value and a predetermined significance level (alpha, typically $\alpha = 0.05$). The test's null hypothesis (H0) states that homoscedasticity is present (i.e., the variance of the errors is constant).

In our scenario, the p-value (0.085) is greater than the standard alpha level of 0.05. Consequently, we must **fail to reject the null hypothesis** of the test. This outcome suggests that there is insufficient statistical evidence to conclude that heteroscedasticity is present within the original regression model. If the p-value had been less than 0.05, we would reject H0 and conclude that heteroscedasticity is a significant issue.

Since we failed to reject the null, the assumption of homoscedasticity holds true for this data, and it is safe to interpret the standard errors of the coefficient estimates in the primary regression summary table without concern for severe bias.

Remedial Actions if Heteroscedasticity is Detected

If the Breusch-Pagan test indicates a rejection of the null hypothesis ($p < 0.05$), this means significant heteroscedasticity is present in the data. When this occurs, the OLS estimates of the coefficients remain unbiased, but the reported standard errors are unreliable, invalidating t-tests and overall model inference. Several common techniques can be employed to correct this issue:

1. Transform the Response Variable.

A frequent and effective approach is to apply a mathematical transformation to the response variable (the dependent variable). For instance, using the logarithm (log) of the response variable instead of its original scale often stabilizes the variance and effectively mitigates heteroscedasticity. Alternatively, using the square root of the response variable is another common variance-stabilizing transformation.

2. Employ Robust Standard Errors.

A more modern and generally preferred approach, which does not require changing the model specification or transforming variables, is the use of Heteroscedasticity-Consistent Standard Errors (HCSEs), often referred to as "robust standard errors." These methods adjust the calculation of the standard errors to account for the non-constant variance without altering the coefficient estimates themselves. While SPSS does not provide the robust standard error option directly in the standard Linear Regression dialogue box, it can be implemented through syntax or specialized add-ons.

3. Use Weighted Regression (Weighted Least Squares).

This specialized type of regression model assigns a specific weight to each data point based on the estimated variance of its error term. Observations associated with higher error variance (where the errors are spread out widely) are given smaller weights, effectively shrinking their influence on the squared residuals and the overall fit. When appropriate weights are accurately estimated, weighted regression can successfully eliminate the problem of heteroscedasticity and restore the reliability of the standard errors.