How to Perform a Breusch-Pagan Test in SPSS to Detect Heteroscedasticity

Authored by

stats writer

January 24, 2026

The Breusch-Pagan test is a crucial statistical tool employed to detect the presence of heteroscedasticity, which translates to unequal variance in the error terms of a dataset. This condition violates a core assumption of standard Ordinary Least Squares (OLS) regression, making the test essential for validating results in regression analysis. When performed using specialized software like SPSS, the process involves running an auxiliary regression on the squared residuals. While some modern statistical packages offer direct options, performing the Breusch-Pagan test in SPSS typically requires a manual, multi-step approach involving the initial regression setup, computation of squared residuals, and a subsequent auxiliary regression run. The detailed procedure ensures accurate identification of variance issues, thereby safeguarding the reliability of your coefficient standard errors.

The generalized steps for executing this procedure are outlined below, providing context for the comprehensive example that follows:

Prepare the dataset in SPSS and define the dependent and independent variables for the primary regression model.

Execute the primary regression analysis, ensuring that the unstandardized residuals are saved as a new variable in the dataset.

Transform the saved residual variable by computing its squared values, resulting in the critical variable used for the auxiliary test.

Perform a second (auxiliary) linear regression, using the squared residuals as the new dependent variable and the original predictors as independent variables.

Analyze the results presented in the output window, specifically focusing on the F-statistic or chi-square value and the corresponding p-value from the ANOVA table to draw conclusions about the presence of heteroscedasticity.

By carefully following these systematic instructions, you can accurately assess whether heteroscedasticity is present in your data, which is paramount for drawing valid inferences from your regression analysis results.

Performing the Breusch-Pagan Test in SPSS: A Comprehensive Guide

The Breusch-Pagan test is fundamentally utilized to assess whether the variance of the errors in a regression model is constant across all levels of the independent variables. This is the definition of **homoscedasticity**, and the test specifically looks for the deviation from this ideal state, known as heteroscedasticity. Identifying this issue is crucial because its presence leads to biased standard errors, rendering the traditional t-statistics and F-statistics unreliable for hypothesis testing, even though the coefficient estimates themselves remain unbiased.

The following detailed, step-by-step guide illustrates precisely how to execute the Breusch-Pagan Test within the SPSS environment, utilizing a practical academic example that requires careful data preparation and manipulation before the final statistical inference.

1. Conceptualizing the Regression Model and Data Preparation

Before launching into the statistical procedure, it is essential to define the context of the analysis. We will employ a practical scenario where we aim to predict a student's final examination score based on their study habits. This requires fitting a **multiple linear regression** model that incorporates key independent variables. Understanding the model structure is the first step toward validating its assumptions.

For this comprehensive example, we hypothesize a model that uses the number of hours spent studying (**hours**) and the number of preparatory exams taken (**prep_exams**) to accurately predict the final exam score of twenty sampled students. The theoretical relationship can be expressed by the following equation, where β represents the coefficient estimates and ε represents the error term:

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

The initial practical step requires inputting the raw data into SPSS. This dataset includes the dependent variable, **score**, and the two independent variables. Ensuring the data is correctly structured, labeled, and assigned the appropriate measurement level in the **Variable View** is critical for the subsequent analysis, preventing potential errors in the regression setup.

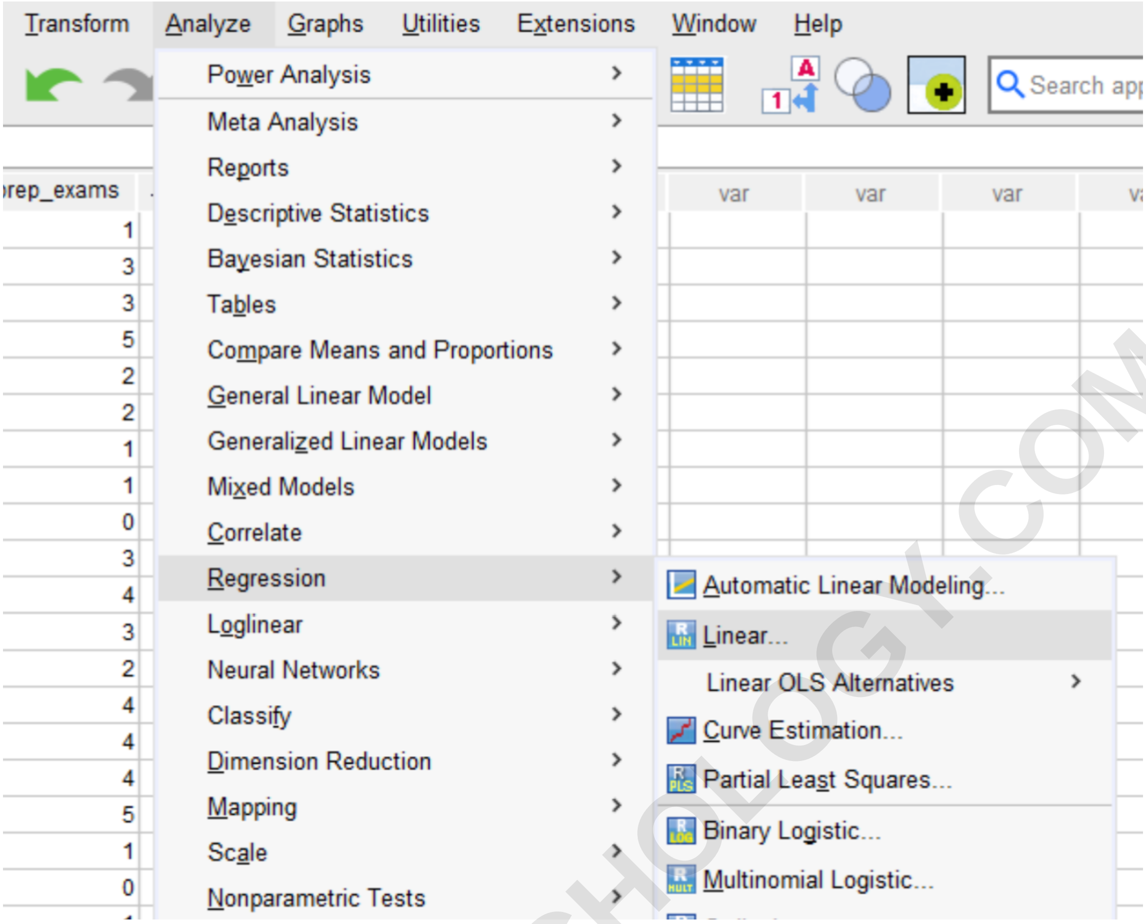
	hours	prep_exams	score	var	var
1	1	1	76		
2	2	3	78		
3	2	3	85		
4	4	5	88		
5	2	2	72		
6	1	2	69		
7	5	1	94		
8	4	1	94		
9	2	0	88		
10	4	3	92		
11	4	4	90		
12	3	3	75		
13	6	2	90		
14	5	4	90		
15	3	4	82		
16	4	4	85		
17	6	5	90		
18	2	1	83		
19	1	0	62		
20	2	1	76		
21					
22					
23					

Once the data entry is complete and verified, we are prepared to move forward with the estimation of the primary regression model. This foundational step generates the error terms, or residuals, which are indispensable for conducting the auxiliary regression required by the Breusch-Pagan test.

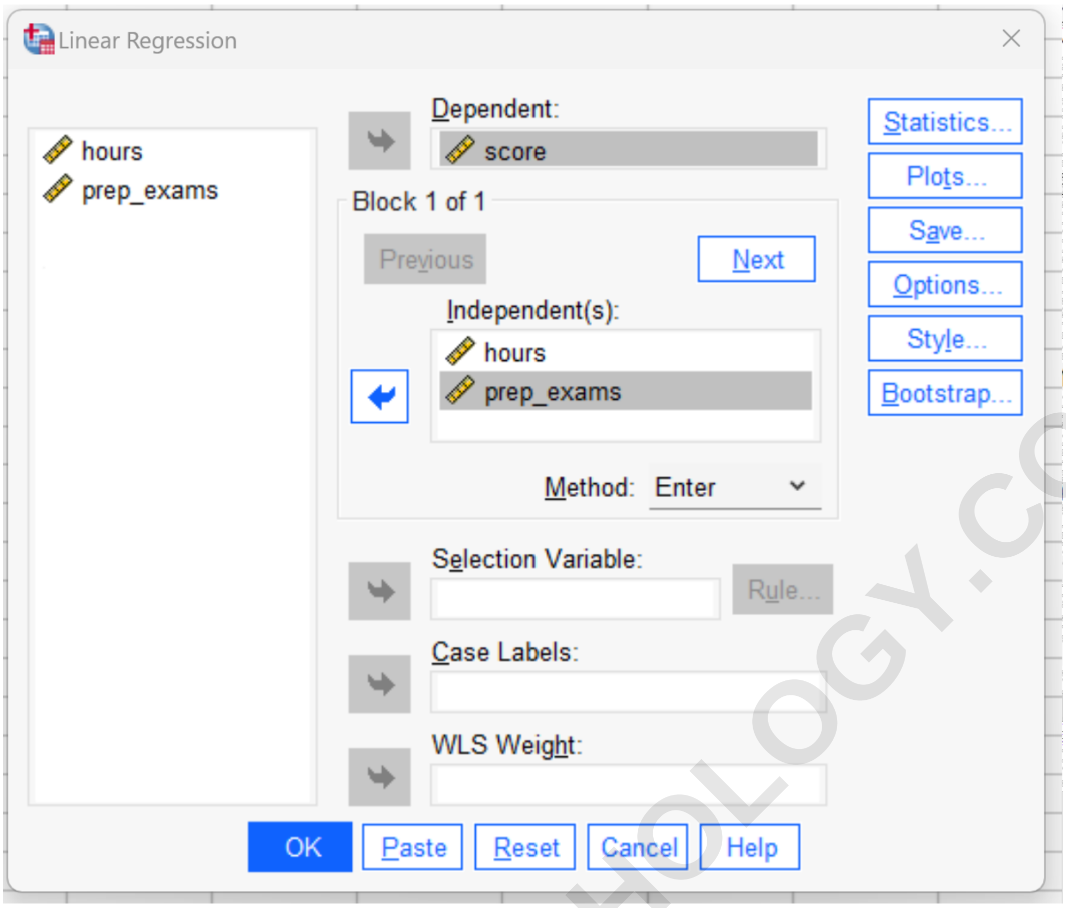
2. Executing the Primary Multiple Linear Regression Model

The next procedural step involves fitting the chosen multiple linear regression model to the dataset. This action provides us with the preliminary estimates of the coefficients and, crucially for the diagnostic phase, the unstandardized residuals. These residuals must be explicitly saved into the dataset for further calculations.

To initiate the regression routine in SPSS, navigate through the menu options: Click the **Analyze** tab, hover over **Regression**, and then select **Linear**. This action opens the main dialogue box where the model's components must be specified accurately according to the research hypothesis.



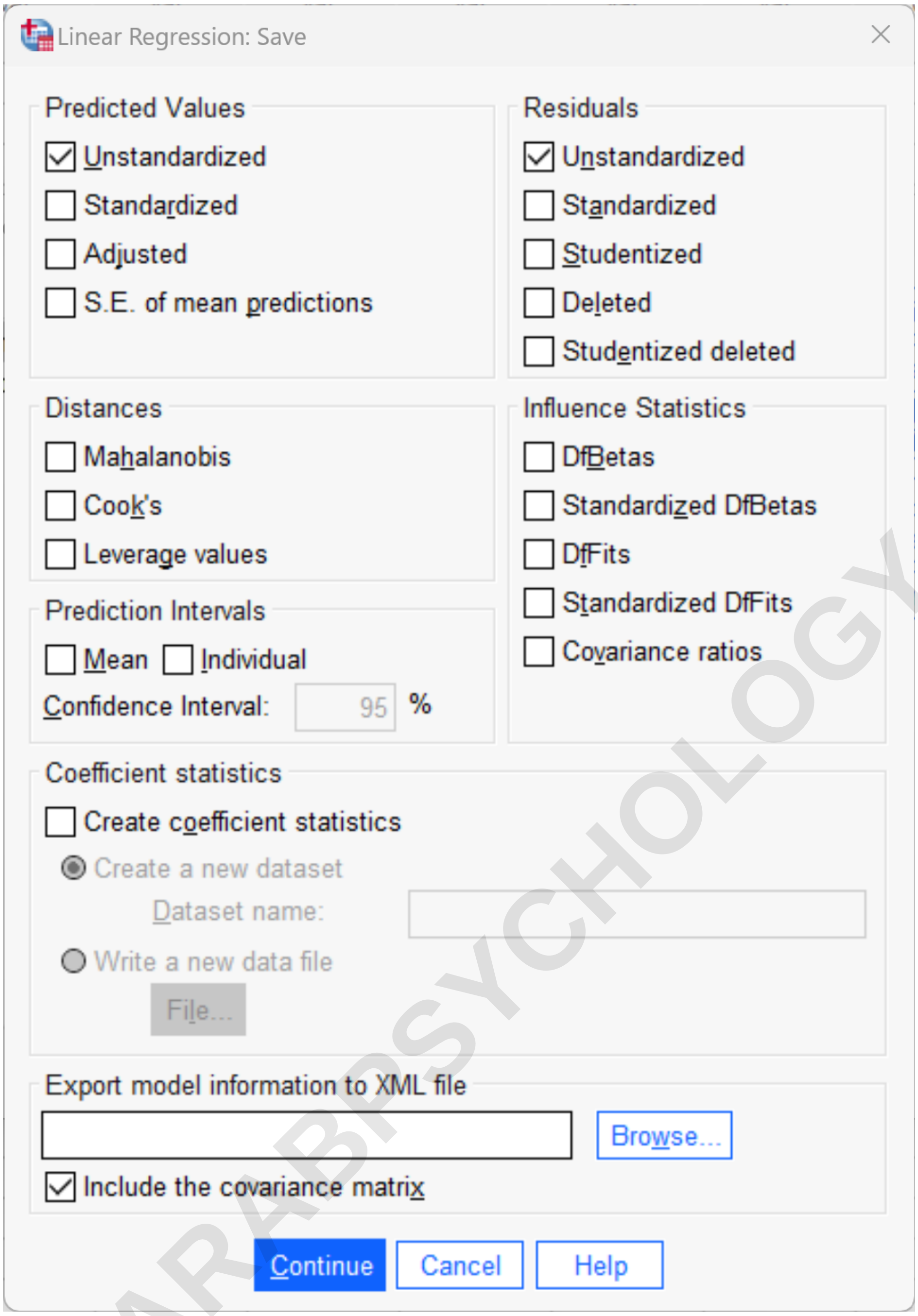
Within the **Linear Regression** dialogue box, carefully assign the variables: drag the dependent variable, **score**, to the **Dependent** panel. Subsequently, drag the predictors, **hours** and **prep_exams**, into the **Independent(s)** panel. It is important to confirm that the variables are correctly specified as this primary regression provides the context for the heteroscedasticity test.



3. Saving Necessary Diagnostic Variables

To proceed with the Breusch-Pagan test manual procedure in SPSS, we must instruct the software to compute and save the residuals generated during the regression process. These residuals represent the unexplained variance in the model and are the foundation for testing the assumption of constant error variance. Click the **Save** button located in the Linear Regression dialogue box to access the options for generating diagnostic variables.

In the new **Save** window, locate the **Predicted Values** section and check the box next to **Unstandardized**. Similarly, locate the **Residuals** section and check the box next to **Unstandardized**. While saving predicted values is often useful for plotting, saving the **Unstandardized Residuals** (typically saved as **RES_1**) is absolutely mandatory for calculating the squared terms needed in the next step.



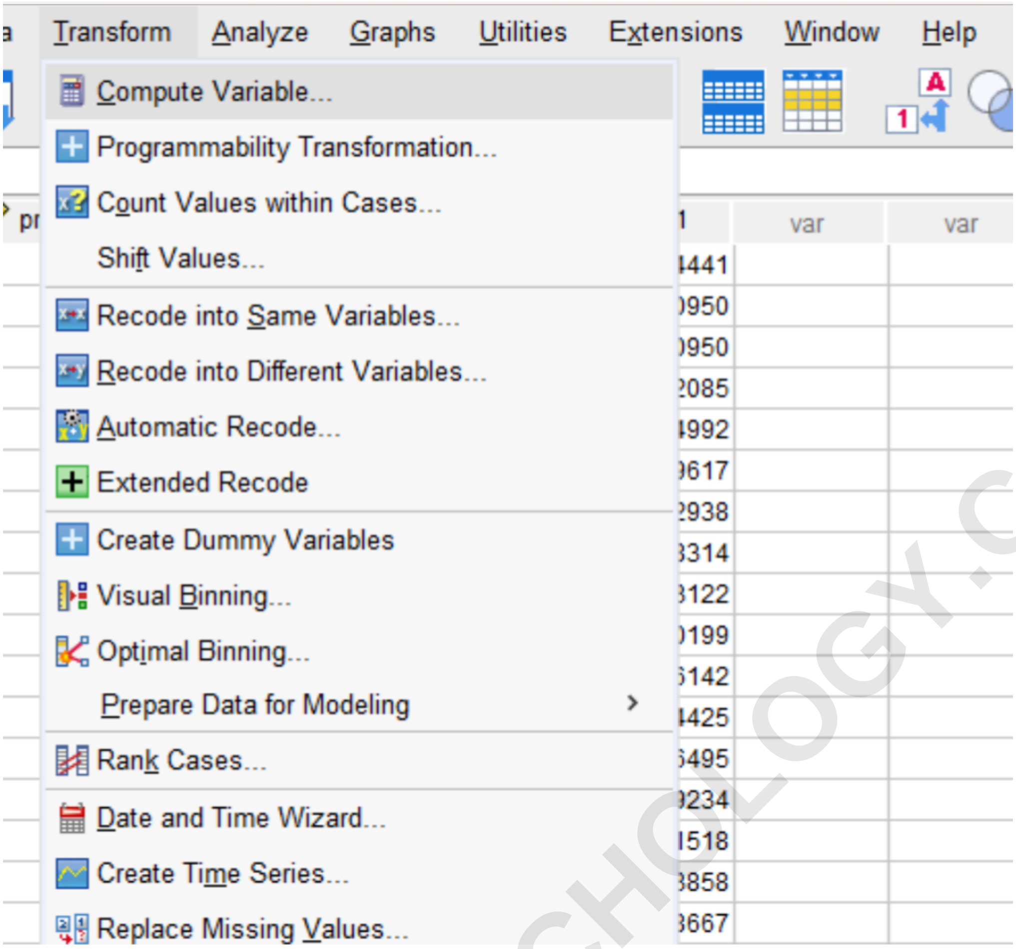
After confirming the selections, click **Continue** to return to the main dialogue box. Finally, click **OK** to execute the regression. SPSS will generate the standard regression output and, crucially, a new variable named **RES_1** in your Data View. We will focus solely on this residual variable for the next stage of the test, ignoring the primary output tables for the moment.

	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					

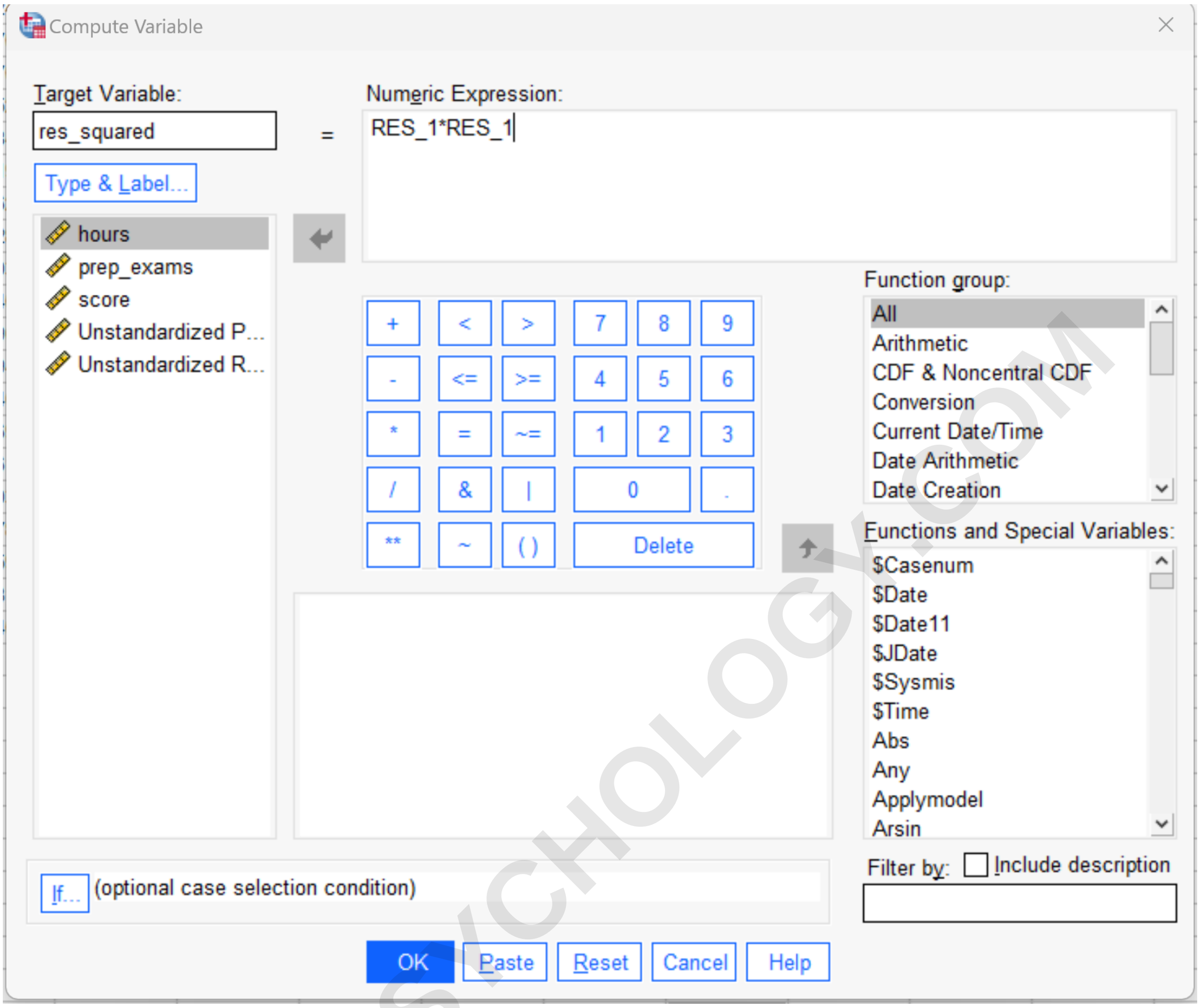
4. Calculating the Squared Residuals for the Test

The theoretical basis of the Breusch-Pagan test is to run an auxiliary regression where the variance of the error term is the dependent variable. Since we cannot directly observe this variance, we utilize the squared residuals as a robust proxy for the variance of the errors across observations. Therefore, the next critical step is to generate a new variable consisting of the squared values of the unstandardized residuals (**RES_1**).

To calculate this new variable, navigate to the **Transform** tab in the SPSS menu and select **Compute Variable**. This feature allows for simple arithmetic manipulation of existing variables to create new data columns, which is essential for transforming the linear residuals into a variance proxy.



In the **Compute Variable** dialogue box, define the parameters for the calculation. First, specify a clear and descriptive **Target Variable** name, such as **res_squared**. Second, define the calculation in the **Numeric Expression** box. Since we saved the unstandardized residuals as **RES_1**, the correct expression to square this variable is **RES_1*RES_1**. This formula instructs SPSS to compute the square of every residual observation in the dataset.



Click **OK** to execute the calculation. Upon returning to the Data View, a new variable named **res_squared** will have been created. This column, holding the squared residuals, is the critical dependent variable for the auxiliary regression that formally conducts the Breusch-Pagan test.

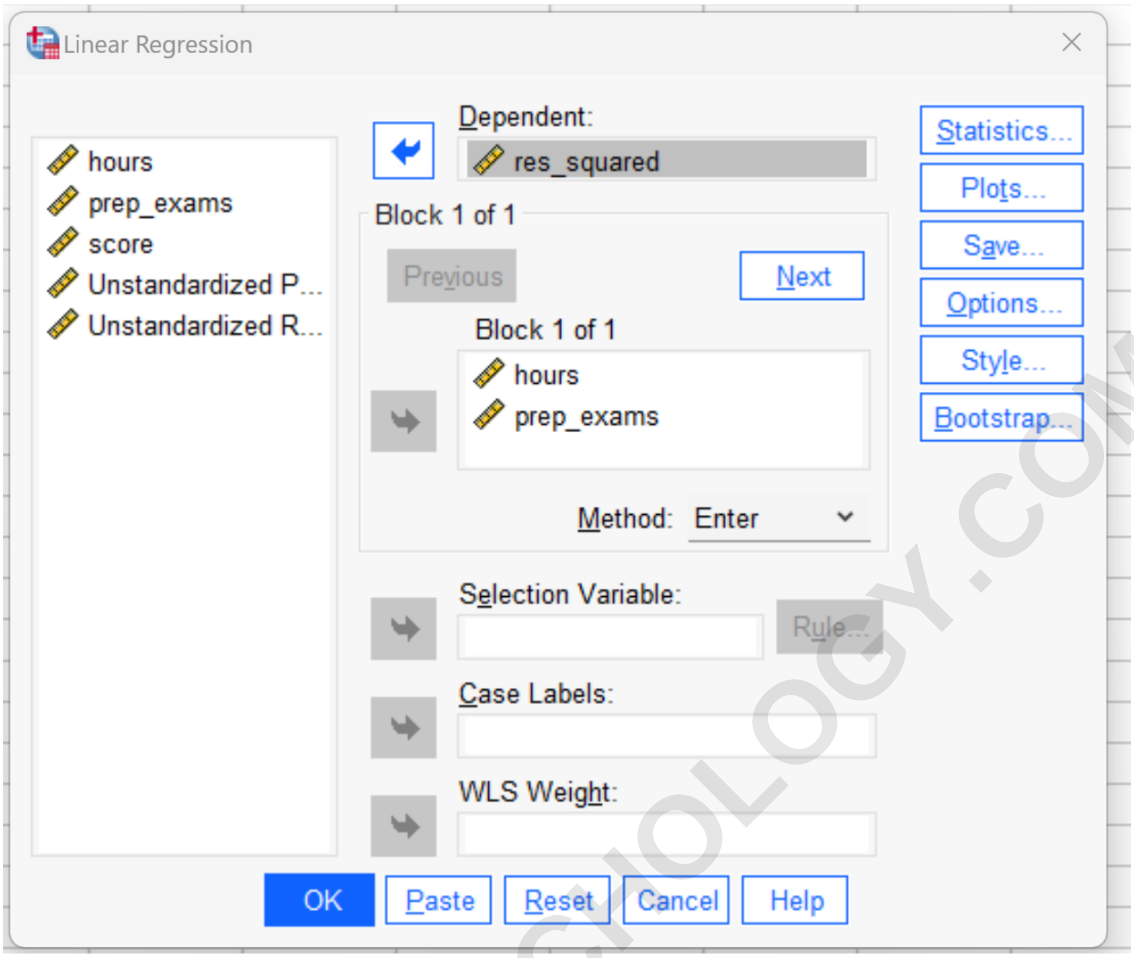
	hours	prep_exams	score	PRE_1	RES_1	res_squared
1	1	1	76	73.75559	2.24441	5.04
2	2	3	78	77.29050	.70950	.50
3	2	3	85	77.29050	7.70950	59.44
4	4	5	88	85.47915	2.52085	6.35
5	2	2	72	77.84992	-5.84992	34.22
6	1	2	69	73.19617	-4.19617	17.61
7	5	1	94	92.37062	1.62938	2.65
8	4	1	94	87.71686	6.28314	39.48
9	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						
23						

5. Conducting the Auxiliary Regression Test

The ultimate step in the Breusch-Pagan test procedure involves running a secondary, or auxiliary, linear regression. The objective is to determine whether the magnitude of the squared residuals (our variance proxy) is systematically predicted by the original independent variables. A statistically significant relationship here is evidence of heteroscedasticity.

To set up this critical auxiliary regression, repeat the initial regression steps: Click **Analyze**, then **Regression**, and finally **Linear**. Crucially, we must now redefine the dependent variable. Drag the newly created variable, **res_squared**, into the **Dependent** panel, replacing the original dependent variable (**score**).

The independent variables, **hours** and **prep_exams**, must remain in the **Independent(s)** panel. This setup formalizes the test of the null hypothesis: that the original predictors have no explanatory power over the variance of the error terms, meaning the error variance is constant.



Before clicking **OK**, ensure you clear any selections made in the **Save** submenu from the previous step, as saving residuals from this auxiliary regression is unnecessary. Click **OK** to execute the auxiliary model. SPSS will now provide a set of output tables; the key to the Breusch-Pagan test lies within the ANOVA table of this new output.

6. Interpreting the Breusch-Pagan Test Results

The interpretation of the Breusch-Pagan test relies entirely on the statistical significance of the auxiliary regression model. We must examine the **F-statistic** and its associated p-value, which is presented in the **ANOVA** table of the auxiliary regression output. The larger the R-squared from this auxiliary regression, the stronger the evidence against homoscedasticity.

The output table will display the results, and the critical value is the significance level located in the **Sig.** column of the **ANOVA** table. This p-value determines whether we reject or fail to reject the null hypothesis (**H0**), which posits that **homoscedasticity** is present (i.e., the variance is constant).

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

In the results shown above, we observe that the p-value (under the **Sig.** column) is **0.085**. Using the conventional significance threshold of α **= 0.05**, we compare the calculated p-value to the threshold. Since 0.085 is greater than 0.05, we must fail to reject the null hypothesis of constant variance.

Failing to reject **H0** leads to the conclusion that we do not have sufficient statistical evidence to claim that heteroscedasticity is present in the original regression model. Therefore, the standard errors of the coefficient estimates derived from the primary regression are considered unbiased and reliable, allowing us to proceed confidently with the interpretation of the coefficient significance.

7. Addressing Heteroscedasticity: Remedial Actions

The outcome of the Breusch-Pagan test guides the subsequent analysis pipeline. If, as in our example, the test suggests that homoscedasticity holds (p-value > 0.05), no immediate remedial action regarding variance is necessary. However, if the test yields a statistically significant result (p-value ≤ 0.05), it confirms the presence of heteroscedasticity, necessitating corrective measures to ensure valid statistical inference.

When heteroscedasticity is confirmed, the standard errors reported by the Ordinary Least Squares (OLS) method are biased, potentially leading to incorrect conclusions about the significance of the independent variables. Researchers typically turn to one of the following methods to address this issue:

Transforming the Response Variable: Applying a mathematical transformation to the dependent variable can often stabilize the variance. The most common and effective technique involves using the **natural logarithm (log)** of the response variable. Log transformation often compresses the spread of larger values, making the variance more consistent across the range of the predictors. Alternatives include the **square root transformation** or the inverse transformation, chosen based on the pattern of variance observed in diagnostic plots.

Implementing Weighted Least Squares (WLS) Regression: WLS is a sophisticated alternative estimation method tailored for data exhibiting known or estimable heteroscedasticity. This method assigns varying weights to each observation in the dataset, inversely proportional to the estimated variance of its error term. Data points associated with higher error variance are given smaller weights, thereby reducing their influence on the parameter estimates. Successfully implementing WLS requires accurately estimating the weights, but when done correctly, it yields efficient and unbiased coefficient estimates and reliable standard errors.

Utilizing Robust Standard Errors: Also referred to as **Huber-White standard errors**, this method is a pragmatic approach that does not attempt to eliminate the heteroscedasticity itself. Instead, it adjusts the computation of the standard errors to provide accurate measures of uncertainty, thus ensuring valid hypothesis testing even in the presence of unequal variance. While SPSS may require advanced modules or specific syntax for implementing robust standard errors, this option is widely favored in econometric and complex statistical modeling because it bypasses the need for variable transformation while preserving the interpretation of the original coefficient estimates.