Breusch-Pagan Test in SPSS: A Step-by-Step Guide to Testing for Heteroscedasticity

Authored by

Mohammed looti

November 12, 2025

Understanding the Breusch-Pagan Test and Homoscedasticity

The Breusch-Pagan Test is an indispensable diagnostic tool specifically designed for rigorous regression analysis. Its fundamental objective is to rigorously assess the presence of non-constant error variance, known as heteroscedasticity, within a statistical model. When analysts employ the standard **Ordinary Least Squares (OLS)** technique for fitting a regression model, a core assumption must be met: the variance of the residuals (or error terms) must remain constant across all observations and levels of the independent variables. This critical assumption is termed **homoscedasticity**.

If this prerequisite of constant variance is violated--meaning the spread of the residuals systematically changes as the predictor variables increase or decrease--the resulting statistical issue is defined as heteroscedasticity. Although the coefficient estimates derived from the OLS procedure remain **unbiased** even when this issue is present, the calculation of their underlying variances becomes significantly flawed. Consequently, critical statistical inferences, such as conducting hypothesis testing or constructing reliable confidence intervals, may be misleading because the calculated standard errors are inaccurate.

To ensure the trustworthiness of your quantitative findings, diagnosing and addressing this assumption violation is paramount. The following comprehensive, step-by-step tutorial provides precise instructions on how to execute the powerful Breusch-Pagan Test (B-P Test) using the robust statistical capabilities of **IBM SPSS Statistics** software. This methodology is essential for validating the foundational assumptions of your regression analysis and securing the reliability of your final results.

Prerequisite: Setting Up Your Data in SPSS

To walk through this demonstration, we will analyze a practical scenario concerning student academic performance. Our objective is to construct a multiple linear regression model capable of predicting a student's final exam score based on two crucial predictor variables: the total number of hours spent studying, and the quantity of preparatory exams taken. Conceptually, the model we are fitting can be formally expressed using the following mathematical notation:

Exam Score = $\beta 0 + \beta 1$(hours) + $\beta 2$(prep exams)

The initial and crucial step before any analysis begins is the accurate input and organization of the raw dataset within the SPSS environment. This example dataset comprises observations for 20 distinct students, meticulously tracking their study hours, the count of prep exams completed, and their corresponding final exam score. Proper organization is critical: each defined variable must be assigned to its own dedicated column within the data structure.

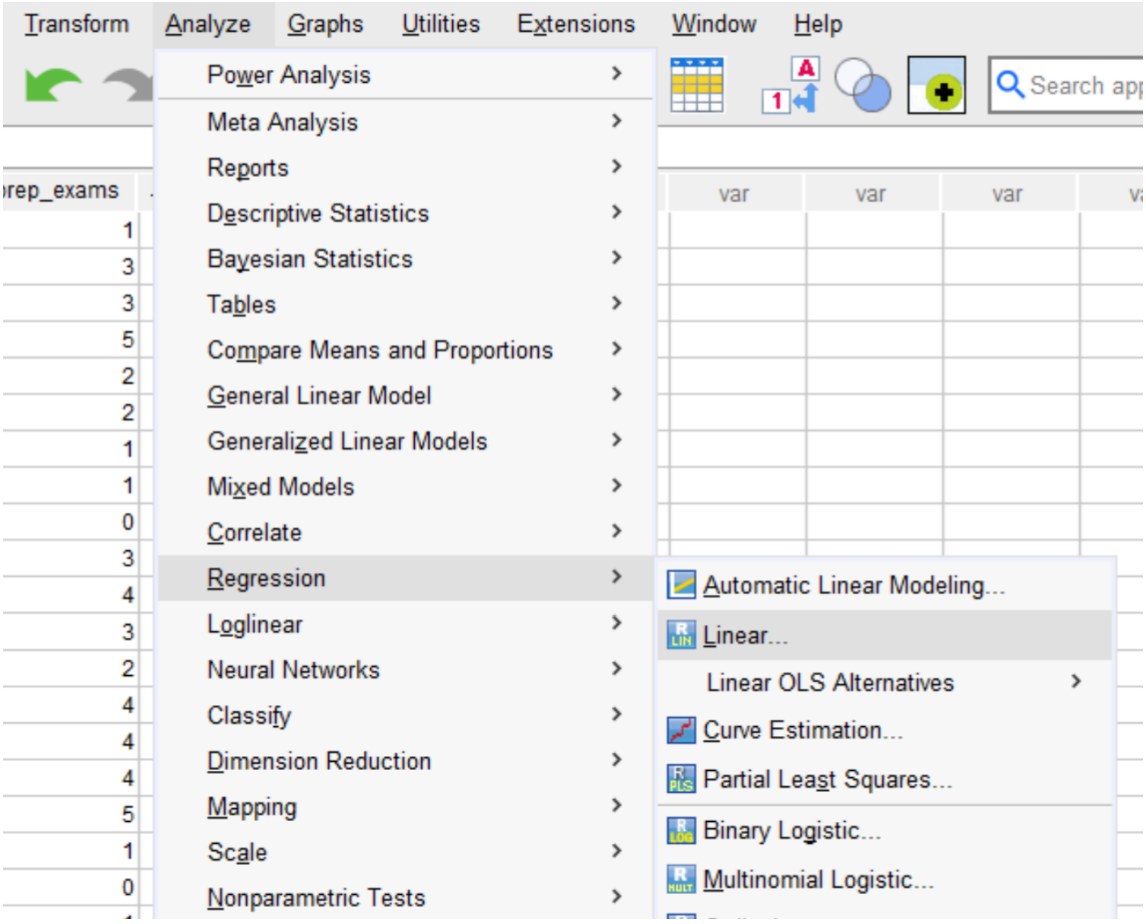
Once the variables (specifically **score**, **hours**, and **prep_exams**) have been correctly defined in the Variable View and the corresponding numerical data has been entered into the Data View window of SPSS, the structure should precisely match the appearance illustrated in the image provided below. Establishing this clean data foundation is the necessary precursor to fitting the primary regression equation and subsequently testing for consistency in variance.

	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					

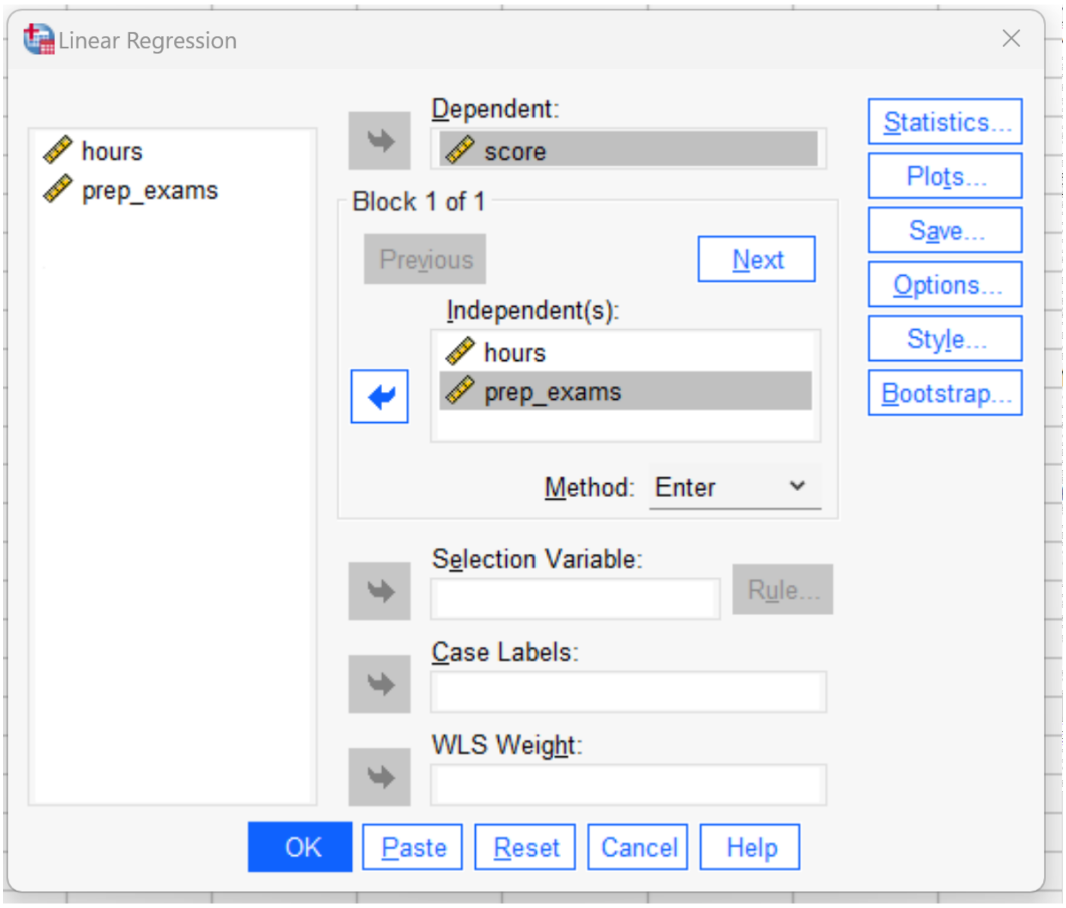
Fitting the Initial Regression Model and Saving Residuals

The next pivotal stage involves executing the primary multiple linear regression model. This process serves a dual purpose: it yields the desired coefficient estimates and, more importantly for the B-P test, it generates the required components known as the residuals. These residuals fundamentally quantify the vertical distance, or difference, between the observed outcome values in the dataset and the predicted values calculated by the fitted model.

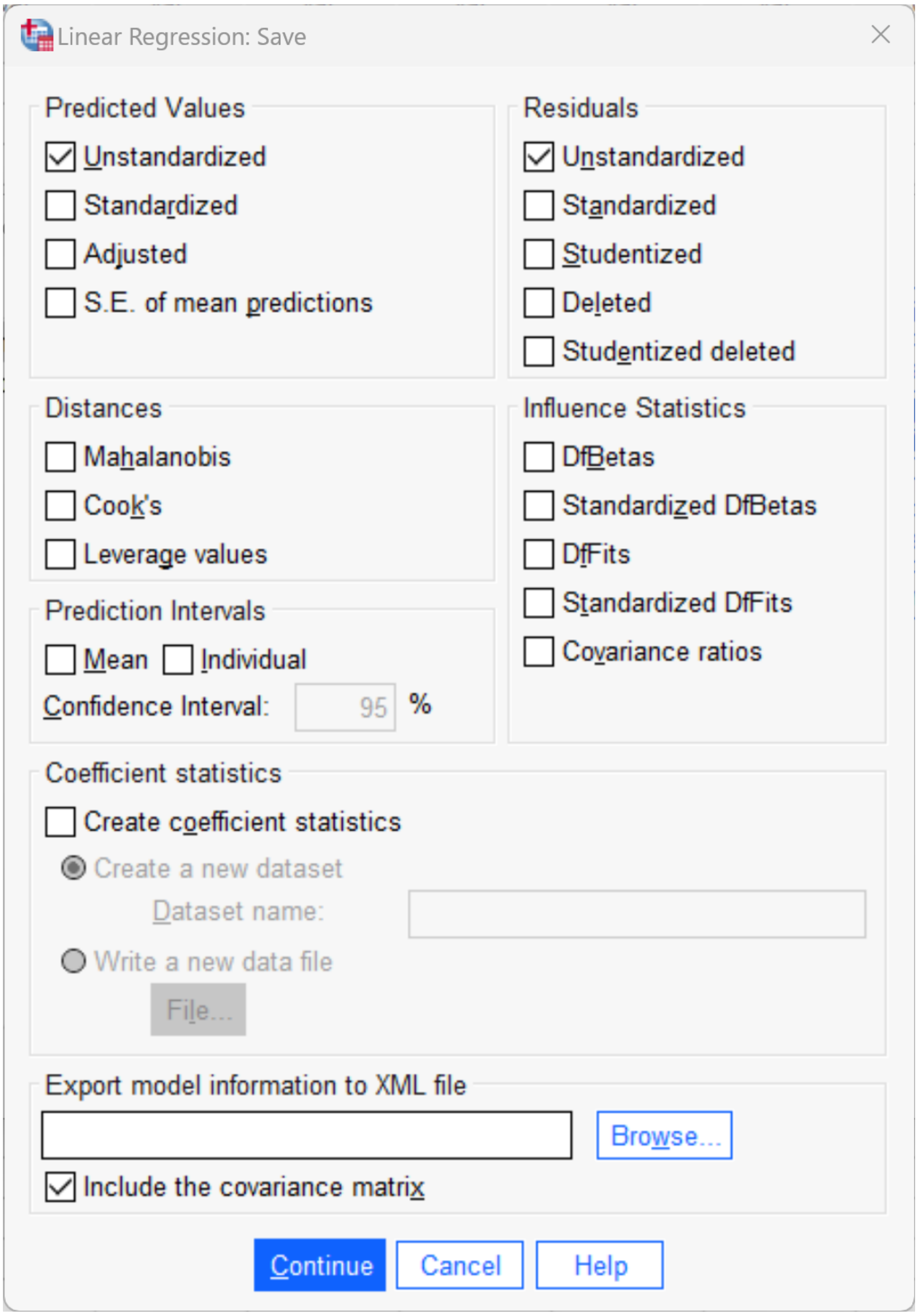
To initiate the model fitting procedure in SPSS, navigate through the main menu using the following structured sequence: start by clicking the **Analyze** tab, hover your cursor over **Regression**, and then select the **Linear** option. This action will immediately launch the Linear Regression dialog box, which is the interface used to precisely specify the model's dependent and independent variables.



Within the Linear Regression window, you must correctly specify the model structure: drag the **score** variable into the **Dependent** variable panel. Subsequently, drag the **hours** and **prep_exams** variables into the **Independent(s)** panel. Once the variables are defined, it is essential to instruct SPSS to compute and save specific output values that are mandatory for the subsequent B-P testing procedure. Locate and click the **Save** button found within the Linear Regression dialog box.



In the resulting Save dialog box, two specific options must be activated to generate the required data points for the test. Under the **Predicted Values** section, ensure the box next to **Unstandardized** is checked. Crucially, under the **Residuals** section, also check the box next to **Unstandardized**. These unstandardized residuals, which SPSS typically saves as a new variable named RES_1, form the indispensable foundation for the B-P test calculation methodology.



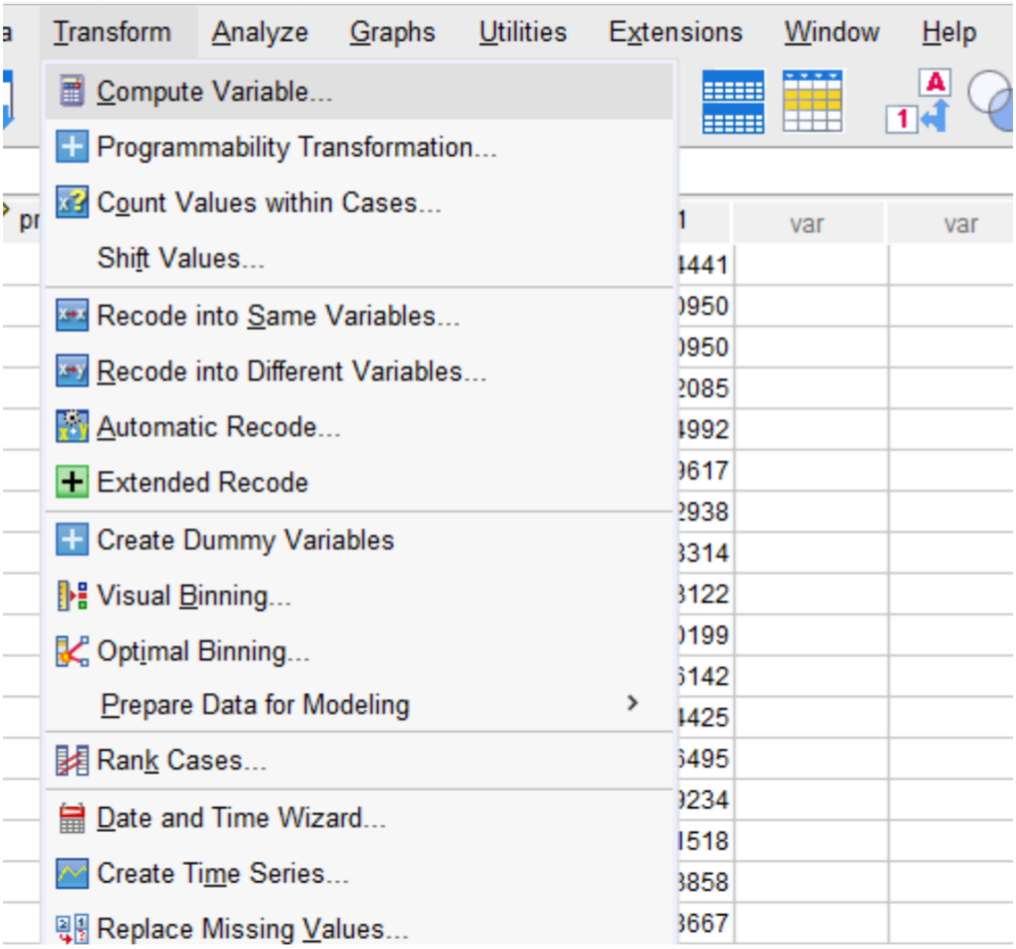
After confirming these selections, click the **Continue** button to close the Save dialog, and then click **OK** in the main Linear Regression window to execute the model computation. SPSS will produce the standard output tables, but, more importantly, two new data columns (representing predicted values and residuals, often named PRE_1 and RES_1) will be automatically appended to your working dataset in the Data View, signifying the completion of the first major step.

	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					

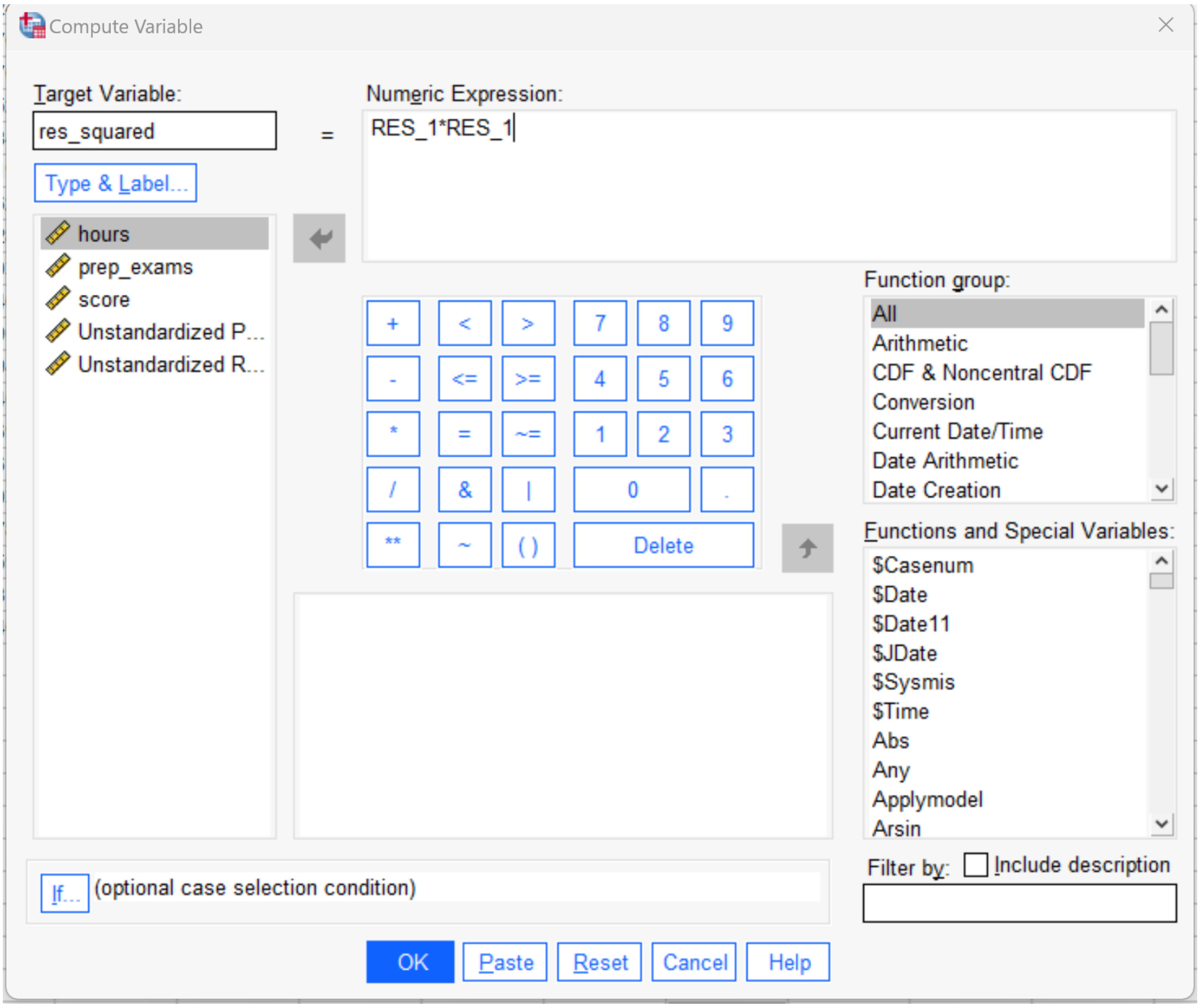
The Mechanism: Transforming Residuals into Squared Errors

The central statistical principle underlying the Breusch-Pagan Test requires us to determine if the variance of the error terms is systematically correlated with the original predictor variables. To quantify this potential relationship, the next operational step is to mathematically create a new variable that holds the squared value of the model's residuals. This derived variable, the squared residual, will then be utilized as the dependent variable in a subsequent, necessary auxiliary regression model.

To initiate this crucial data transformation, click the **Transform** tab located in the SPSS menu bar, and then choose the **Compute Variable** option. This robust feature is designed specifically to allow users to generate new variables based on complex mathematical operations applied directly to existing data columns within the dataset.



In the subsequent Compute Variable dialog box, you must first define the output variable by specifying its name in the **Target Variable** field--we will use **res_squared** for clarity and convention. Next, accurately define the calculation required in the **Numeric Expression** box. Given that the unstandardized residuals were saved automatically as RES_1 in the preceding step, the correct formula required to obtain the squared residuals is simply **RES_1*RES_1**. This formula systematically calculates the squared error term for every single observation in the dataset.



Once **OK** is clicked, SPSS executes this mathematical calculation instantly, generating the new variable, **res_squared**. This column is then appended to the existing dataset structure and contains the squared error terms, values which are directly proportional to the estimated variance. This transformation successfully concludes all the necessary preparatory work required before performing the final inferential stage of the Breusch-Pagan Test.

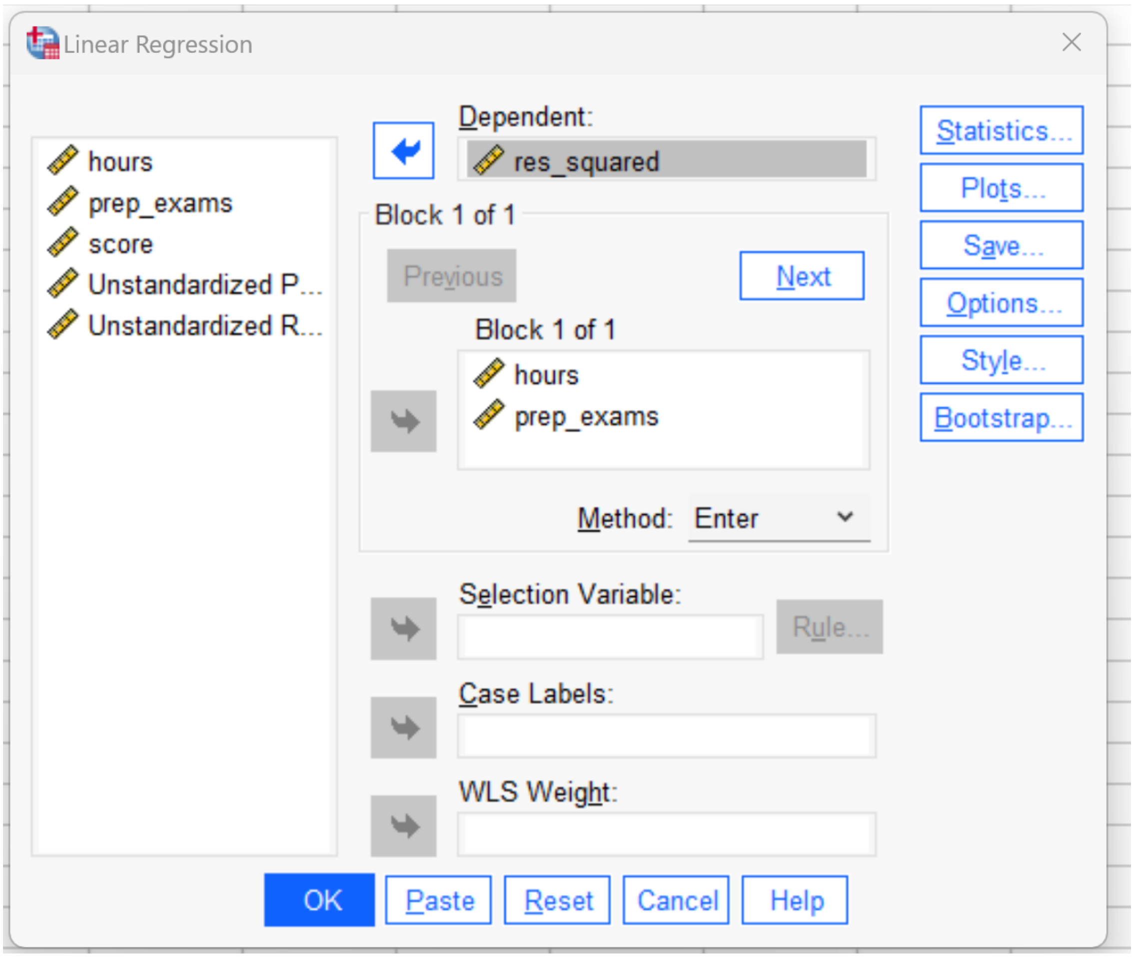
	hours	prep_exams	score	PRE_1	RES_1	res_squared
1	1	1	76	73.75559	2.24441	5.04
2	2	3	78	77.29050	.70950	.50
3	2	3	85	77.29050	7.70950	59.44
4	4	5	88	85.47915	2.52085	6.35
5	2	2	72	77.84992	-5.84992	34.22
6	1	2	69	73.19617	-4.19617	17.61
7	5	1	94	92.37062	1.62938	2.65
8	4	1	94	87.71686	6.28314	39.48
9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

Executing the Breusch-Pagan Test Regression

The inferential component of the B-P Test is accomplished by running a second, specialized auxiliary regression. In this new model, the squared residuals (our newly created **res_squared** variable) are regressed onto the set of original independent variables (in our case, **hours** and **prep_exams**). The logic is powerful: if the original independent variables possess significant predictive power over the squared residuals, it strongly implies that the variance of the errors is not constant, which definitively signals the presence of heteroscedasticity.

To perform this auxiliary regression, navigate back to the primary regression dialog box by clicking **Analyze**, then **Regression**, and finally selecting **Linear**. It is essential at this juncture to redefine the model parameters from the previous step.

In the Linear Regression dialog box, replace the initial dependent variable (score) with the newly computed dependent variable, **res_squared**. Ensure that the original predictor variables, **hours** and **prep_exams**, remain correctly placed in the **Independent(s)** panel. Importantly, because our sole interest lies in the overall fit statistics of this specific auxiliary model, there is no requirement to click the **Save** button again or adjust any other settings within the dialog box.



Click **OK** to execute the auxiliary regression model. SPSS will rapidly generate a distinct set of output tables pertaining solely to this model. Although the classic B-P test statistic is traditionally calculated based on the R-squared value from this output, in the standard SPSS presentation, the most relevant inferential statistic--the p-value--is conveniently provided within the **ANOVA** output table, which is automatically generated by the procedure.

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

Interpreting the Breusch-Pagan Test Results

The final interpretation of the Breusch-Pagan Test hinges entirely on the p-value extracted from the auxiliary regression's ANOVA table. The test is constructed around a standard framework of statistical hypothesis testing, defining two mutually exclusive hypotheses:

Null Hypothesis (H0): The model exhibits **homoscedasticity**; that is, the variance of the error terms is constant and stable.

Alternative Hypothesis (Ha): Heteroscedasticity is present; the variance of the errors is non-constant and varies systematically with the predictor variables.

The crucial p-value required for making the statistical decision is located in the **Sig.** column of the **ANOVA** table for the auxiliary regression. Observing the sample output provided, the p-value for the model is calculated to be **.085**.

The statistical decision rule is applied straightforwardly: if the calculated p-value is less than the pre-determined significance level (alpha, conventionally set at 0.05), we possess sufficient evidence to **reject the null hypothesis**. Conversely, if the p-value exceeds 0.05, we **fail to reject the null hypothesis**. Since our calculated p-value of 0.085 is greater than the 0.05 threshold, we must fail to reject the null hypothesis. This finding leads to the conclusive statistical determination that we lack sufficient evidence to assert that heteroscedasticity is present in the original regression model. Consequently, the fundamental assumption of homoscedasticity is deemed valid for this specific dataset.

This conclusion represents a favorable outcome, as it provides confirmation that the calculated standard errors, and thus all subsequent t-tests and confidence intervals derived from the original regression summary table, are statistically reliable and trustworthy. Analysts can confidently proceed with the interpretation of the coefficient estimates without concern for potentially biased error estimates.

Addressing Heteroscedasticity: Robust Remedial Actions

If the outcome had been different--for instance, if the Breusch-Pagan Test had resulted in the rejection of the null hypothesis (p-value < 0.05)--it would unequivocally signal that heteroscedasticity is indeed present. In such challenging scenarios, the standard errors reported in the regression output are compromised and unreliable, necessitating immediate corrective action to ensure the validity of any statistical inferences.

Fortunately, statisticians have developed several robust methods specifically designed to address or effectively mitigate the detrimental effects of non-constant variance. The selection of the most appropriate method typically hinges on a careful assessment of the severity and the underlying functional form of the observed heteroscedasticity:

Transformation of the Response Variable: This is often the most straightforward and highly effective strategy. It involves applying a mathematical transformation, such as the natural logarithm (log), the square root, or the inverse function, to the dependent (response) variable instead of using the raw score. This technique frequently stabilizes the variance structure, causing the symptoms of heteroscedasticity to disappear.

Using Heteroscedasticity-Consistent Standard Errors (HCSE): A popular and elegant alternative is the use of HCSE methods, such as **White's Correction**. This approach is highly valued because it maintains the original functional form of the model and avoids transforming the variables. Instead, HCSE procedures adjust the underlying calculation of the standard errors to accurately account for the non-constant variance, yielding robust standard errors without altering the primary coefficient estimates.

Weighted Least Squares (WLS) Regression: This is an advanced technique that fundamentally modifies the OLS estimation process. WLS assigns differential weights to each observation in the dataset, with the weight inversely proportional to the estimated variance of that observation's error term. Data points associated with higher variance (which typically produce larger residuals) receive comparatively smaller weights, effectively reducing their undue influence on the overall model fit and solving the problem of non-constant variance. Implementing WLS generally requires an accurate estimation of the variance structure, which adds a layer of complexity.

By selecting and applying the most suitable remedial measure, the analyst ensures that the regression analysis remains statistically rigorous and sound, thereby enabling the reliable extraction of accurate conclusions regarding the causal or correlational relationships between the predictors and the outcome variable.