

# How do I perform a Breusch-Pagan Test in SAS?

Authored by  
**stats writer**

November 19, 2025

## RECOMMENDED CITATION

stats writer (2025). *How do I perform a Breusch-Pagan Test in SAS?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=97024>

## Understanding the Breusch-Pagan Test

The Breusch-Pagan Test (BPT) is a foundational statistical test crucial for validating assumptions within a linear regression model. Its primary purpose is to rigorously assess the presence of heteroscedasticity--a condition where the variance of the error terms (residuals) is not constant across all levels of the independent variables. Detecting this issue is paramount because heteroscedasticity violates one of the core assumptions of Ordinary Least Squares (OLS) regression, potentially leading to unreliable standard errors and inaccurate statistical inferences.

When running regression analysis in software like SAS, addressing potential heteroscedasticity is a standard requirement for robust modeling. If the BPT indicates that the variance is non-constant, the standard errors calculated for the coefficients will be biased, meaning subsequent hypothesis tests and confidence intervals cannot be trusted. Therefore, learning how to correctly implement and interpret the BPT in SAS is an essential skill for any quantitative analyst.

## Setting Up the Breusch-Pagan Test in SAS

While the traditional Breusch-Pagan Test can be executed using the standard `PROC REG` and the `TEST` statement, modern SAS environments, especially when dealing with complex econometric models, often utilize the powerful `PROC MODEL` procedure. This procedure provides greater flexibility for fitting complex equations and explicitly includes the required statement for the BPT. This tutorial focuses on the `PROC MODEL` approach, which streamlines the process of fitting the linear regression model and checking the diagnostic assumption simultaneously.

To begin, we must first establish a properly structured dataset and define the functional form of the model we intend to test. The process involves creating input variables and a response variable, followed by specifying the model equation within the `PROC MODEL` block. The subsequent steps will demonstrate how to integrate the test request directly into the model fitting procedure, ensuring the model's validity against the assumption of homoscedasticity.

## Example Scenario: Predicting Exam Scores

To illustrate the implementation of the Breusch-Pagan Test, consider a common scenario in educational statistics: predicting student performance. We aim to fit a multiple linear regression model where the final exam score is predicted by two factors: the number of hours spent studying and the number of preparatory exams taken. The conceptual model is structured as follows:

$$\text{Exam Score} = \beta_0 + \beta_1(\text{Hours}) + \beta_2(\text{Prep Exams}) + \varepsilon$$

Before fitting the model, we must first generate the sample data. The following SAS code block utilizes the `DATA` step to create a dataset named `exam_data` containing 20 observations, recording

the hours studied, prep exams taken, and the resulting score for each student. This step ensures that the data is correctly ingested into the SAS environment for the subsequent modeling procedure. We then use `PROC PRINT` to display the newly created dataset, confirming its structure and contents.

```
/*create dataset for regression analysis*/
```

```
data exam_data;
```

```
input hours prep_exams score;
```

```
datalines;
```

```
1 1 76
```

```
2 3 78
```

```
2 3 85
```

```
4 5 88
```

```
2 2 72
```

```
1 2 69
```

```
5 1 94
```

```
4 1 94
```

```
2 0 88
```

```
4 3 92
```

```
4 4 90
```

```
3 3 75
```

```
6 2 90
```

```
5 4 90
```

```
3 4 82
```

```
4 4 85
```

```
6 5 90
```

```
2 1 83
```

```
1 0 62
```

```
2 1 76
```

```
;
```

```
run;
```

```
/*view dataset structure and contents*/
```

```
proc print data=exam_data;
```

The resulting dataset structure, as displayed by the `PROC PRINT` statement, confirms the successful creation of the input variables necessary for our regression analysis:

Obs	hours	prep_exams	score
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

## Executing the Breusch-Pagan Test Using PROC MODEL

The next critical step is utilizing the `PROC MODEL` procedure in SAS, which is specialized for estimating parameters in simultaneous equation models, though it is perfectly suited for standard linear regression while enabling advanced diagnostic tests. Within this procedure, we define our model equation and specify the parameters to be estimated ( $a_1, b_1, b_2$ ).

The core of the diagnostic check lies in the `FIT` statement. By including the `PAGAN=` option immediately following the variable to be fitted (`score`), we instruct SAS to perform the Breusch-Pagan Test. The variables listed within the parenthesis--(`1 hours prep_exams`)--represent the variables hypothesized to be related to the variance of the errors. We include the constant term (`1`) and all predictors (`hours` and `prep_exams`) for a comprehensive test.

The `OUT=resid1` and `OUTRESID` options are included to save the residuals and fitted values into a new dataset, which is often useful for subsequent manual diagnostic plots, although they are not strictly necessary for running the BPT itself. The complete code block for fitting the model and

executing the test is shown below:

```
/*fit regression model and perform Breusch Pagan test for heteroscedasticity*/
proc model data=exam_data;
parms a1 b1 b2;
score = a1 + b1*hours + b2*prep_exams;
fit score / pagan=(1 hours prep_exams)
out=resid1 outresid;
run;
quit;
```

## Interpreting the SAS Output

After executing the code, SAS generates extensive output related to the model estimation. We must scroll to the diagnostic section, specifically the table titled "Breusch-Pagan Test," to find the results relevant to our hypothesis test. This section summarizes the calculation derived from regressing the squared residuals onto the independent variables. The image below highlights the key output table containing the final result of the test.

The MODEL Procedure							
Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
score	3	17	585.1	34.4198	5.8668	0.6158	0.5706

  

Nonlinear OLS Parameter Estimates				
Parameter	Estimate	Approx Std Err	t Value	Approx Pr >  t
a1	69.66126	3.0788	22.63	<.0001
b1	4.653756	0.9832	4.73	0.0002
b2	-0.55943	0.9998	-0.56	0.5831

  

Number of Observations		Statistics for System	
Used	20	Objective	29.2568
Missing	0	Objective*N	585.1369

  

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
score	Breusch-Pagan	5.05	2	0.0803	1, hours, prep_exams

The output table provides two critical values for the test: the chi-square test statistic and the associated p-value. In this particular run, the test statistic is calculated as **5.05**, and the corresponding p-value is **0.0803**. These two figures are used to make the final determination regarding the presence of heteroscedasticity.

## Drawing Conclusions from the Breusch-Pagan Test

To interpret these results, we must recall the structure of the Breusch-Pagan Test. The BPT operates under the following framework:

The null hypothesis ( $H_0$ ) states that the variance of the errors is constant across all observations (i.e., homoscedasticity is present). Conversely, the alternative hypothesis ( $H_a$ ) states that the variance is not constant (i.e., heteroscedasticity is present).

We typically compare the calculated p-value against a predetermined significance level, often  $\alpha = 0.05$ . If the p-value is less than  $\alpha$ , we reject the null hypothesis, concluding that heteroscedasticity exists. In our example, the calculated p-value of **0.0803** is greater than 0.05. Therefore, we **fail to reject the null hypothesis**.

The statistical conclusion is that we do not have sufficient evidence, at the 5% significance level, to claim that heteroscedasticity is present in our linear regression model. Because the assumption of homoscedasticity holds, we can safely proceed to interpret the standard errors and coefficient estimates provided in the standard regression summary table, knowing that they are reliable and unbiased.

## Addressing Significant Heteroscedasticity

While our example demonstrated a case where the assumption of homoscedasticity was maintained, it is critical for analysts to know the proper remedial steps should they reject the null hypothesis. A significant BPT result confirms the presence of heteroscedasticity, rendering the standard errors derived from the OLS model unreliable, even though the coefficient estimates themselves remain unbiased.

When unreliable standard errors are present, any hypothesis testing or confidence interval construction based on the regression output is flawed. Fortunately, statisticians have developed several robust methods to address this issue, ensuring that the model conclusions are trustworthy. The choice of method often depends on the severity and nature of the non-constant variance observed.

The most common strategies for mitigating heteroscedasticity include:

**Transforming the Response Variable:** Applying a mathematical function to the dependent

variable can often stabilize the variance of the residuals. For instance, using the natural logarithm (log transformation) of the response variable instead of the original raw values is a highly effective, standard practice that frequently resolves heteroscedasticity issues. Alternatively, using the square root transformation is also a viable option, particularly for count data.

**Using Weighted Regression (WLS):** Weighted Least Squares regression assigns differential weights to each observation based on the estimated variance of its error term. Data points associated with higher variance (which contribute more to the heteroscedasticity problem) receive smaller weights, thereby reducing their influence on the estimation of the regression coefficients. If the weights are correctly specified to be inversely proportional to the error variance, the issue of non-constant variance is mathematically eliminated, resulting in efficient and unbiased coefficient estimates and reliable standard errors.

**Employing Robust Standard Errors:** A simpler alternative is to use heteroscedasticity-consistent standard errors, such as White's standard errors (or HC standard errors). This method does not attempt to fix the heteroscedasticity itself but rather provides corrected standard errors that are robust to its presence, allowing for reliable inference without transforming the model or weighting the observations. SAS procedures like `PROC REG` and `PROC GLM` often support options (e.g., `COVB(HC3)`) to calculate these robust error estimates.

## Conclusion: Ensuring Model Validity

The ability to correctly perform and interpret the Breusch-Pagan Test in SAS is fundamental to rigorous statistical modeling. By using the specialized `PROC MODEL` procedure and the `PAGAN` option, analysts can efficiently verify the homoscedasticity assumption. When the test yields non-significant results, as in our example, confidence in the model's inferential statistics is maintained. Should the test prove significant, employing techniques like variable transformation or Weighted Least Squares ensures that the final model remains statistically valid and reliable for prediction and inference.