

Performing the Breusch-Pagan Test for Heteroscedasticity in SAS: A Step-by-Step Guide

Authored by
Mohammed Iooti

November 14, 2025

RECOMMENDED CITATION

Mohammed Iooti (2025). *Performing the Breusch-Pagan Test for Heteroscedasticity in SAS: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=1605>

Achieving statistically sound results in [regression analysis](#), particularly when applying the widely used Ordinary Least Squares (OLS) estimation method, rests upon several fundamental econometric assumptions. Chief among these is the assumption of **homoscedasticity**. This critical requirement demands that the variance of the model's error terms must remain uniform and constant across all observations and levels of the predictor variables. When this assumption is violated--a common occurrence in real-world data--the model suffers from [heteroscedasticity](#). This statistical affliction renders the calculated standard errors biased and unreliable, potentially leading to erroneous inferences regarding the true statistical significance of the independent variables. To formally diagnose and quantify this detrimental presence of non-constant variance, analysts rely on the powerful statistical technique known as the [Breusch-Pagan Test](#).

The [Breusch-Pagan Test](#) is specifically designed as a rigorous diagnostic procedure to detect the systemic relationship between the variability of the residuals and the independent variables incorporated into the regression framework. Its methodology involves an auxiliary regression where the squared residuals--representing the unexplained error--are regressed on the predictors. A statistically significant result from this test acts as a definitive warning, indicating that the foundational assumption of homoscedasticity has been violated. Such a finding necessitates immediate remedial intervention to ensure the robustness and accuracy of the regression findings before they can be utilized for critical forecasting, policy setting, or business decision-making processes.

This authoritative guide offers expert, step-by-step instruction on how to effectively execute the [Breusch-Pagan Test](#) within [SAS](#), one of the premier statistical software environments utilized extensively by professional statisticians and researchers globally. We will meticulously navigate every phase of the process, starting with the necessary data preparation and structuring, proceeding through the actual model fitting using specialized SAS procedures, and culminating in a comprehensive interpretation of the test results. To facilitate seamless practical application, this tutorial includes precise explanations complemented by ready-to-use [SAS](#) code examples.

The Foundation of Reliable Regression: The Homoscedasticity Assumption

Prior to delving into the technical mechanics of implementing the Breusch-Pagan Test, it is essential to establish a firm theoretical understanding of the consequences of [heteroscedasticity](#) and why its presence fundamentally compromises the integrity of [regression analysis](#). In an ideal statistical setting characterized by homoscedasticity, the spread or dispersion of the model's residuals remains uniform and consistent across the entire spectrum of predicted values. This visual constancy signifies that the model's errors are equally predictable everywhere.

Conversely, when [heteroscedasticity](#) infects the data, the residual spread fluctuates dramatically. This variability often manifests as a noticeable widening or narrowing of the residual scatter as the

values of the independent variables change, meaning the error terms lack a constant variance. This erratic behavior directly violates a primary assumption required for the efficiency and statistical validity of OLS estimation, creating profound complications for the subsequent inferential statistics derived from the model.

The most devastating consequence of non-constant error variance is the distortion of the [standard errors](#) associated with the [regression coefficients](#). While OLS retains its property of producing unbiased point estimates for the coefficients even under heteroscedasticity, the calculated standard errors become unreliable, potentially being severely underestimated or overestimated. This inaccuracy critically undermines the construction of confidence intervals and the reliability of hypothesis testing. For instance, if the [standard errors](#) are erroneously compressed, an analyst might incorrectly conclude that a predictor variable is statistically significant when, in reality, it is not. Such profound errors in inference can lead to misdirected conclusions and flawed decisions based on the statistical model's output.

Practical Implementation in SAS: Setting Up the Case Study

To provide a clear and actionable demonstration of applying the Breusch-Pagan Test within [SAS](#), we will utilize a realistic example scenario. Consider a research objective focused on predicting the final examination scores of students based on two crucial explanatory metrics: the total number of hours dedicated to studying and the quantity of preparatory exams completed throughout the course. We hypothesize that the relationship between these variables can be effectively modeled using a multiple linear regression structure, formally represented as:

$$\text{Exam Score} = \beta_0 + \beta_1(\text{hours}) + \beta_2(\text{prep exams}) + \varepsilon$$

In this formulation, β_0 represents the model's intercept, while β_1 and β_2 are the estimated [coefficients](#) for study hours and preparatory exams, respectively. The term ε denotes the stochastic error term. Our primary analytical goal is twofold: first, to accurately fit this model to the empirical data, and second, to critically assess whether the crucial assumption of homoscedasticity holds true for the error term ε across all observations.

The initial and indispensable step for any statistical procedure in [SAS](#) is the precise creation and definition of the dataset. Below, we construct a sample dataset, named **exam_data**, comprising observations for 20 hypothetical students. This dataset includes measurements for the final scores, hours studied, and the number of preparatory exams taken. This foundational SAS code snippet is crucial for loading the required data into the statistical environment, preparing it for subsequent model estimation and rigorous diagnostic testing procedures.

```
/*create dataset*/  
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*/
```

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

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 Formal Diagnostic Test Using PROC MODEL

Once the dataset is successfully established within the [SAS](#) environment, the subsequent critical phase involves the simultaneous fitting of the regression model and the execution of the [Breusch-Pagan Test](#). This complex task is handled with high efficiency and flexibility by [PROC MODEL](#), a versatile procedure within SAS designed for the estimation and testing of linear and nonlinear systems, including essential diagnostic checks for underlying model assumptions.

The syntax employed by [PROC MODEL](#) is both structured and intuitive. We begin by defining the parameters of our regression equation using the `parms` statement, followed by the definition of the model where `score` is specified as the dependent variable. The core command that triggers the diagnostic test is embedded within the `fit` statement. By including the option `pagan=(1 hours prep_exams)`, we explicitly instruct SAS to compute the Breusch-Pagan Test statistic. The variables listed within the parentheses--specifically `1` (representing the intercept), `hours`, and `prep_exams`--are the predictors utilized in the auxiliary regression, which models the squared

residuals against these variables. This auxiliary step is the methodological engine of the Breusch-Pagan procedure, testing for any direct relationship between the residual variance and the independent variables.

Furthermore, the `out=resid1` `outresid` options, also specified within the `fit` statement, fulfill a practical secondary objective: they instruct SAS to calculate and save the model's residuals into a newly generated dataset named `resid1`. This dataset of residuals is invaluable for subsequent visual diagnostic checks, such as generating scatter plots of residuals against predicted values. Upon execution, the following SAS code will produce the complete regression output, prominently featuring the crucial results required for our formal analysis of the Breusch-Pagan Test.

```
/*fit regression model and perform Breusch Pagan test*/  
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;
```

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

Interpreting the Breusch-Pagan Test Results

Following the successful execution of the SAS code, the output generated by [PROC MODEL](#) will include various tables summarizing the regression fit and diagnostic checks. For the specific goal of assessing variance stability, the most critical section is the table dedicated to the Breusch-Pagan Test results. This section provides the essential numerical evidence required to formally determine whether the assumption of variance constancy, or homoscedasticity, is tenable for our [regression model](#).

From the presented output, we must identify two fundamental values: the calculated [test statistic](#) and its corresponding [p-value](#). In the results generated by our case study example, the test statistic is calculated as **5.05**, and the associated p-value is determined to be **0.0803**. The interpretation of these statistics is governed by the established null and alternative hypotheses inherent to the Breusch-Pagan Test:

Null Hypothesis (H0): The model exhibits homoscedasticity, implying the variance of the error terms is constant.

Alternative Hypothesis (H1): The model exhibits [heteroscedasticity](#), implying the variance of

the error terms is non-constant.

To finalize the decision regarding the [null hypothesis](#), we compare the calculated [p-value](#) to a pre-defined significance level (α), which is conventionally set at 0.05. The decision rule is unambiguous: if the p-value is less than α , we must reject the null hypothesis. Conversely, if the p-value is equal to or greater than α , we fail to reject the [null hypothesis](#). Since our computed p-value of 0.0803 is clearly greater than 0.05, we conclude that we fail to reject the null hypothesis.

Failing to reject the [null hypothesis](#) leads to a favorable conclusion: we lack statistically sufficient evidence to confidently assert that [heteroscedasticity](#) is present within our regression model. This outcome validates that the assumption of homoscedasticity is tenable for this specific dataset. Consequently, we can proceed with confidence, interpreting the [standard errors](#) of the [coefficient estimates](#) as reliable, thereby ensuring the integrity of the inferences drawn from the regression summary table concerning the statistical significance of our predictor variables.

Addressing Violations: Strategies for Correcting Heteroscedasticity

Although our specific case study yielded a favorable result where non-constant variance was not statistically confirmed, it is vital for any competent analyst to understand how to proceed should the Breusch-Pagan Test produce a significant result (i.e., rejection of the [null hypothesis](#)). If the test statistically indicates the presence of [heteroscedasticity](#), the calculated [standard errors](#) of the [coefficient estimates](#) become unreliable, potentially skewing conclusions about the true statistical importance of the predictors.

Fortunately, statistical methodology provides several established and highly effective strategies for mitigating or correcting the problems introduced by unstable variance. These remedies are designed to restore the consistency and efficiency of the residuals' variance, thereby safeguarding the validity of the regression model's inferences. Two of the most commonly accepted and utilized methods are detailed below:

Applying a [Transformation](#) to the Response Variable. A practical and often effective solution involves applying a mathematical [transformation](#) to the dependent variable. Common transformations include taking the logarithm (log) or the square root of the variable's original values. The logarithmic transformation is frequently the first line of defense, as it inherently compresses larger values more drastically than smaller ones. This compression can effectively stabilize the variance, particularly in cases where the magnitude of the variance is observed to increase proportionally with the mean of the dependent variable, thus successfully promoting homoscedasticity. Similarly, the square root transformation is useful in contexts where the error variance is believed to be proportional to the mean itself.

Utilizing [Weighted Least Squares \(WLS\) Regression](#). An alternative and statistically robust

technique involves adopting [weighted regression](#). Unlike standard OLS, which incorrectly assumes that all observations contribute equally to the variance structure, WLS regression assigns a specific weight to each data point. These weights are calculated to be inversely proportional to the estimated variance of the error term for that observation. This weighting scheme ensures that data points associated with higher variances--which are the primary source of the problem in [heteroscedasticity](#)--are given smaller weights, thereby diminishing their undue influence on the estimation process. By appropriately adjusting the influence of each observation, WLS provides more efficient and reliable [coefficient estimates](#) and accurate [standard errors](#), directly correcting the core issues caused by non-constant variance.

The choice between these corrective approaches should be guided by the specific nature of the data, the underlying theoretical model, and the pattern of [heteroscedasticity](#) observed in diagnostic residual plots. It is considered best practice to re-examine residual plots after implementing any corrective measure to confirm that the issue has been successfully resolved.

Conclusion: Validating Model Reliability with Formal Diagnostics

The [Breusch-Pagan Test](#) is justly regarded as an indispensable and fundamental component of the diagnostic toolkit for any statistical practitioner engaged in rigorous [regression analysis](#). Its specialized capacity to formally and statistically evaluate the presence of non-constant error variance is essential for guaranteeing the validity, efficiency, and ultimate reliability of all subsequent regression model inferences. As demonstrated meticulously through our guided example using [SAS](#), the implementation of this test is straightforward, relying on the versatile capabilities of [PROC MODEL](#) and its dedicated `pagan` option.

Interpreting the [p-value](#) derived from the test is paramount for making sound statistical judgments. A non-significant result, such as the one observed in our case study, provides powerful reassurance that the homoscedasticity assumption is met, allowing for confident interpretation of [standard errors](#) and statistical significance. Conversely, a significant result requires immediate analytical attention, signaling a violation that necessitates the application of corrective strategies. These strategies, such as data [transformation](#) or the utilization of [weighted regression](#), are crucial steps for counteracting the biases introduced in the coefficient estimates and their standard errors.

By diligently integrating formal diagnostic procedures like the Breusch-Pagan Test into the standard modeling workflow and promptly applying appropriate solutions when variances are found to be unstable, researchers and analysts can construct more robust and credible statistical models. This rigorous practice significantly enhances the integrity of statistical findings, leading directly to more accurate predictive modeling and facilitating better-informed, evidence-based decision-making across a diverse range of academic and industrial applications.