

How can I perform a Breusch-Pagan test in SAS?

Authored by
stats writer

June 23, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I perform a Breusch-Pagan test in SAS?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=148414>

The Breusch-Pagan test, also known as the White test, is a statistical method used to determine the presence of heteroscedasticity in a regression model. To perform this test in SAS, one must first run the desired regression model using the PROC REG procedure. Then, the user can use the SAS macro %BPTEST to obtain the test statistic and p-value for the Breusch-Pagan test. This macro automatically adjusts for the number of independent variables in the model and provides a simple and efficient way to conduct this test in SAS. Additionally, SAS also offers other procedures such as PROC GLM and PROC MIXED that can be used to perform the Breusch-Pagan test. Overall, SAS provides various options for users to easily and accurately perform the Breusch-Pagan test in their analysis.

Perform a Breusch-Pagan Test in SAS

A is used to determine if is present in a regression analysis.

This tutorial explains how to perform a Breusch-Pagan Test in SAS.

Example: Breusch-Pagan Test in SAS

Suppose we want to fit a multiple linear regression model that uses number of hours spent studying and number of prep exams taken to predict the final exam score of students:

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

First, we'll use the following code to create a dataset that contains this information for 20 students:

```
/*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 printdata=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

Next, we'll use `proc model` to fit this multiple linear regression model along with the `pagan` statement to perform the Breusch-Pagan test for heteroscedasticity:

```

/*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

The last table in the output shows the results of the

Breusch-Pagan test.

From this table we can see that the test statistic is 5.05 and the corresponding p-value is 0.0803.

Since the p-value is not less than 0.05, we fail to reject the null hypothesis.

This means we do not have sufficient evidence to say that heteroscedasticity is present in the regression model.

Thus, it's safe to interpret the standard errors of the coefficient estimates in the regression summary table.

What To Do Next

However, if you reject the null hypothesis, this means heteroscedasticity is present in the data. In this case, the standard errors that are shown in the output table of the regression may be unreliable.

There are a couple common ways that you can fix this issue, including:

1. Transform the response variable. You can try performing a transformation on the response variable.

For example, you could use the log of the response variable instead of the original response variable.

Typically taking the log of the response variable is an effective way of making heteroscedasticity go away.

Another common transformation is to use the square root of the response variable.

2. Use weighted regression. This type of regression assigns a weight to each data point based on the variance of its fitted value.

This gives small weights to data points that have higher variances, which shrinks their squared residuals.

When the proper weights are used, this can eliminate the problem of heteroscedasticity.