

How to Test for Homoscedasticity Using the Breusch-Pagan Test in Excel

Authored by
stats writer

December 28, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Test for Homoscedasticity Using the Breusch-Pagan Test in Excel*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=109443>

The Breusch-Pagan test is a crucial diagnostic tool in econometrics and statistics, specifically designed to evaluate one of the core assumptions of classical linear regression: homoscedasticity. This assumption posits that the variance of the error terms (residuals) in a regression model is constant across all levels of the independent variables. When this condition is violated--meaning the variance changes--the model suffers from heteroscedasticity, which compromises the reliability of standard error estimates and statistical inference.

In essence, the Breusch-Pagan test assesses the null hypothesis of homoscedasticity. It operates by calculating a test statistic, often referred to as the Lagrange multiplier statistic, which follows a chi-squared distribution under the null hypothesis. The degrees of freedom for this distribution are determined by the number of independent variables included in the regression model. Understanding how to execute and interpret this test in an accessible platform like Microsoft Excel is vital for validating regression results, ensuring the accuracy of your coefficients, and making reliable predictions based on your data.

This comprehensive guide details the step-by-step process of implementing the Breusch-Pagan test within Excel. We will leverage Excel's built-in statistical functionalities and formulas to move beyond the basic regression output and rigorously test for the presence of variance irregularity, or heteroscedasticity, in your analysis, ensuring all steps are documented for clean and valid results.

Understanding Homoscedasticity and Its Implications

Before diving into the mechanics of the test, it is essential to appreciate why homoscedasticity is such a critical assumption for Ordinary Least Squares (OLS) regression. When the errors are homoscedastic, the OLS estimators are considered the Best Linear Unbiased Estimators (BLUE). However, if the error variance is not constant (i.e., if heteroscedasticity is present), the OLS estimates remain unbiased, but they lose their efficiency. More dangerously, the calculated standard errors will be biased, leading to inaccurate t-statistics and p-values, which can result in incorrect conclusions regarding the statistical significance of the predictors in the model.

The presence of heteroscedasticity often occurs in cross-sectional data, particularly when dealing with variables that exhibit wide ranges of scale, such as income, firm size, or population density. In such cases, the variability of the dependent variable may naturally increase or decrease as the independent variable changes. For instance, predicting expenditure based on income might show that low-income individuals have very similar spending habits (low error variance), while high-income individuals display wildly divergent spending habits (high error variance). The Breusch-Pagan test provides a formal, quantitative method to detect this irregularity, which is often difficult to spot merely through visual inspection of residual plots.

The Statistical Foundation of the Breusch-Pagan Test

The fundamental logic behind the Breusch-Pagan test involves performing an auxiliary regression. Instead of modeling the dependent variable, we model the variance of the residuals. The procedure tests whether the variance of the errors is systematically related to the independent variables. If the variance is truly constant (homoscedastic), then the independent variables should have no explanatory power when regressing them against the variance.

Specifically, the test uses the squared residuals from the original OLS model as the new dependent variable in a secondary, auxiliary regression. The original explanatory variables are used again as predictors. If the explanatory variables significantly predict the squared residuals, it suggests that the variance is changing systematically with the independent variables. The test statistic is derived directly from the R-squared value of this auxiliary regression, making it powerful yet straightforward to calculate, especially using spreadsheet software like Excel.

Example Setup: Analyzing Basketball Player Attributes

To practically demonstrate the application of the Breusch-Pagan test, we will utilize a sample dataset describing the performance attributes of 10 basketball players. Our objective is to fit a model predicting player rating and subsequently test whether the assumption of stable error variance holds true.

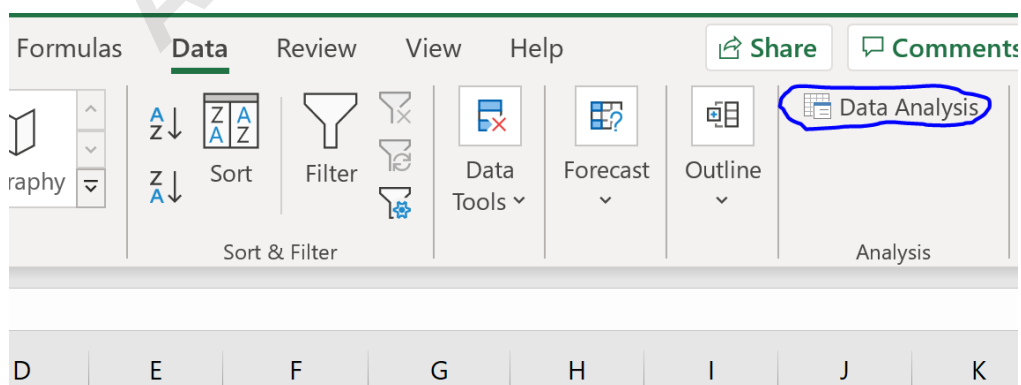
We will construct a multiple linear regression model using **rating** as the response variable (Y) and **points**, **assists**, and **rebounds** as the explanatory variables (X). The Breusch-Pagan test will then be performed to determine if heteroscedasticity is present in the fitted regression. The raw data used for this example is presented below:

	A	B	C	D	E	F
1	rating	points	assists	rebounds		
2	90	25	5	11		
3	85	20	7	8		
4	82	14	7	10		
5	88	16	8	6		
6	94	27	5	6		
7	90	20	7	9		
8	76	12	6	6		
9	75	15	9	10		
10	87	14	9	10		
11	86	19	5	7		
12						
13						
14						
15						
16						
17						
18						

Step 1: Performing the Initial Multiple Linear Regression

The first critical task is generating the necessary output for the primary linear regression model. This step provides us with the estimated coefficients and, most importantly, the residual values, which are the basis for the entire Breusch-Pagan procedure. We utilize the powerful **Data Analysis Toolpak** add-in within Excel to execute this operation.

To access the tool, navigate to the **Data** tab and click on **Data Analysis**. If this option is unavailable, you must first enable the Data Analysis Toolpak add-in. Once the dialog box appears, select *Regression* and click OK. The following window will prompt you to define the ranges for your variables:



Fill in the Input Y Range (Rating) and the Input X Range (Points, Assists, Rebounds). Crucially, ensure that the **Residuals** option is checked in the output section, as these values are essential for the subsequent calculations. After confirming the input ranges and clicking OK, Excel will generate the comprehensive regression summary, providing the coefficients necessary for predicting the response variable:

	A	B	C	D	E	F	G
1	rating	points	assists	rebounds			
2	90	25	5	11			
3	85	20	7	8			
4	82	14	7	10			
5	88	16	8	6			
6	94	27	5	6			
7	90	20	7	9			
8	76	12	6	6			
9	75	15	9	10			
10	87	14	9	10			
11	86	19	5	7			

Regression ? X

Input

Input Y Range: ↑

Input X Range: ↑

Labels Constant is Zero

Confidence Level: %

Output options

Output Range: ↑

New Worksheet Ply:

New Workbook

Residuals

Residuals Residual Plots

Standardized Residuals Line Fit Plots

Normal Probability

Normal Probability Plots

OK Cancel Help

This process results in the standard statistical output table, which summarizes the model's fit and provides the specific coefficient estimates for the intercept and each explanatory variable. We will refer back to these coefficients in the next step to manually calculate the predicted values for each observation:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	rating	points	assists	rebounds										
2	90	25	5	11										
3	85	20	7	8										
4	82	14	7	10										
5	88	16	8	6										
6	94	27	5	6										
7	90	20	7	9										
8	76	12	6	6										
9	75	15	9	10										
10	87	14	9	10										
11	86	19	5	7										
12														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														

Step 2: Calculating Predicted Values and Squared Residuals

To isolate the measure of variance for testing, we must first calculate the predicted response values based on the fitted model and then determine the size of the prediction errors. We begin by calculating the predicted value (\hat{Y}) for each player using the coefficients obtained in Step 1.

The predicted value is calculated by multiplying the coefficient for each explanatory variable by its corresponding observed value and summing the results, along with the intercept term. We must use the exact coefficients from the regression output to ensure accuracy:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	rating	points	assists	rebounds	predicted										
2	90	25	5	11	=H\$22+\$H\$23*B2+\$H\$24*C2+\$H\$25*D2										
3	85	20	7	8											
4	82	14	7	10											
5	88	16	8	6											
6	94	27	5	6											
7	90	20	7	9											
8	76	12	6	6											
9	75	15	9	10											
10	87	14	9	10											
11	86	19	5	7											
12															
13															
14															
15															
16															
17															
18															
19															
20															
21															
22															
23															
24															
25															
26															

When implementing this formula in Excel, it is essential to use absolute cell references for the

coefficients so that the formula can be accurately copied down to calculate the predicted values for all 10 observations. This step transforms the static coefficient estimates into dynamic predicted scores based on the input data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	rating	points	assists	rebounds	predicted										
2	90	25	5	11	90.16632										
3	85	20	7	8	87.6198										
4	82	14	7	10	80.0483										
5	88	16	8	6	84.88144										
6	94	27	5	6	94.54382										
7	90	20	7	9	87.19203										
8	76	12	6	6	78.63733										
9	75	15	9	10	82.93443										
10	87	14	9	10	81.8151										
11	86	19	5	7	85.16144										
12															
13															
14															
15															
16															
17															
18															
19															
20															
21															
22															
23															
24															
25															
26															

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.789024							
R Square	0.622559							
Adjusted R	0.433839							
Standard E	4.584449							
Observatio	10							
ANOVA								
	df	SS	MS	F	ignificance F			
Regression	3	207.996967	69.33232	3.298842	0.099468			
Residual	6	126.103033	21.01717					
Total	9	334.1						
Coefficients Standard Error t Stat P-value Lower 95% Upper 95% Lower 95.0% Upper 95.0%								
Intercept	62.47163	14.5882158	4.282335	0.005192	26.77555	98.16771	26.77555	98.16771
X Variable 1	1.119326	0.41088233	2.724202	0.03445	0.113934	2.124719	0.113934	2.124719
X Variable 2	0.883401	1.38066717	0.639836	0.545919	-2.49497	4.261772	-2.49497	4.261772
X Variable 3	-0.42777	0.85100939	-0.50266	0.633114	-2.51012	1.654574	-2.51012	1.654574

Next, we calculate the squared residuals (\hat{u}^2). The residual is the difference between the actual observed rating and the predicted rating. We square this difference to measure the magnitude of the error while simultaneously ensuring that all values are positive, which is necessary for variance analysis:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	rating	points	assists	rebounds	predicted	squared residuals										
2	90	25	5	11	90.16632	= (E2-A2)^2										
3	85	20	7	8	87.6198											
4	82	14	7	10	80.0483											
5	88	16	8	6	84.88144											
6	94	27	5	6	94.54382											
7	90	20	7	9	87.19203											
8	76	12	6	6	78.63733											
9	75	15	9	10	82.93443											
10	87	14	9	10	81.8151											
11	86	19	5	7	85.16144											
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
22																
23																
24																
25																
26																

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.78902409							
R Square	0.62255901							
Adjusted R	0.43383852							
Standard E	4.58444895							
Observatio	10							
ANOVA								
	df	SS	MS	F	ignificance F			
Regression	3	207.997	69.33232	3.298842	0.099468			
Residual	6	126.103	21.01717					
Total	9	334.1						
Coefficients Standard Error t Stat P-value Lower 95% Upper 95% Lower 95.0% Upper 95.0%								
Intercept	62.4716322	14.58822	4.282335	0.005192	26.77555	98.16771	26.77555	98.16771
X Variable 1	1.11932645	0.410882	2.724202	0.03445	0.113934	2.124719	0.113934	2.124719
X Variable 2	0.88340122	1.380667	0.639836	0.545919	-2.49497	4.261772	-2.49497	4.261772
X Variable 3	-0.42777131	0.851009	-0.50266	0.633114	-2.51012	1.654574	-2.51012	1.654574

Repeating this calculation for all data points generates the full series of squared residuals. This series, which quantifies the unexplained variance at each observation point, will become the new dependent variable for the next stage of the test:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	rating	points	assists	rebounds	predicted	squared residuals										
2	90	25	5	11	90.16632	0.027661										
3	85	20	7	8	87.6198	6.863348										
4	82	14	7	10	80.0483	3.809141										
5	88	16	8	6	84.88144	9.725433										
6	94	27	5	6	94.54382	0.295745										
7	90	20	7	9	87.19203	7.884707										
8	76	12	6	6	78.63733	6.955505										
9	75	15	9	10	82.93443	62.955129										
10	87	14	9	10	81.8151	26.883184										
11	86	19	5	7	85.16144	0.703180										
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
22																
23																
24																
25																
26																

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.78902409							
R Square	0.62255901							
Adjusted R	0.43383852							
Standard E	4.58444895							
Observatio	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	3	207.997	69.33232	3.298842	0.099468			
Residual	6	126.103	21.01717					
Total	9	334.1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	62.4716322	14.58822	4.282335	0.005192	26.77555	98.16771	26.77555	98.16771
X Variable	1.11932645	0.410882	2.724202	0.03445	0.113934	2.124719	0.113934	2.124719
X Variable	0.88340122	1.380667	0.639836	0.545919	-2.49497	4.261772	-2.49497	4.261772
X Variable	-0.42777131	0.851009	-0.50266	0.633114	-2.51012	1.654574	-2.51012	1.654574

Step 3: Performing the Auxiliary Regression

The third step executes the core mechanism of the Breusch-Pagan test: performing a second multiple linear regression. This auxiliary regression uses the calculated squared residuals (\hat{u}^2) as the response variable (Y), while the original explanatory variables (Points, Assists, and Rebounds) remain the predictor variables (X).

If the null hypothesis of homoscedasticity is true, the independent variables should have no power to explain the variance of the squared residuals, meaning the R-squared value (R^2_{new}) of this auxiliary regression should be statistically insignificant. We use the same **Data Analysis Toolpak** process as before, but ensure the Input Y Range now points to the column containing the squared residuals. The output of this auxiliary regression is necessary to extract the critical R^2_{new} value:

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.774852							
R Square	0.600395							
Adjusted R Square	0.400593							
Standard Error	14.96632							
Observations	10							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	3	2019.241	673.0802155	3.004945	0.116639			
Residual	6	1343.945	223.9908692					
Total	9	3363.186						
	<i>Coefficient</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-66.7099	47.62447	-1.400748657	0.210828	-183.243	49.82297	-183.243	49.82297
X Variable 1	0.243934	1.34136	0.181855763	0.861684	-3.03826	3.526125	-3.03826	3.526125
X Variable 2	9.527298	4.507306	2.1137458	0.078963	-1.50168	20.55628	-1.50168	20.55628
X Variable 3	1.216264	2.778193	0.437789653	0.676859	-5.58173	8.014257	-5.58173	8.014257

From this auxiliary regression summary, we extract the crucial R^2_{new} value, which is displayed as 0.600395. This value indicates that approximately 60% of the variance in the squared residuals is being explained by the combination of the original predictor variables. This substantial R-squared value strongly suggests a relationship exists, but we must formally test its statistical significance in the final step.

Step 4: Calculating the Breusch-Pagan Test Statistic

The final quantitative step involves calculating the Breusch-Pagan test statistic, often referred to as the LM statistic. This statistic transforms the R^2 of the auxiliary regression into a value that can be compared against the chi-squared distribution.

The formula for the Breusch-Pagan test statistic (X^2) utilizes the sample size (n) and the R-Squared from the new regression (R^2_{new}):

$$X^2 = n * R^2_{\text{new}}$$

Where:

n: Represents the total number of observations in the dataset.

R^2_{new} : Is the R Square value obtained from the auxiliary regression (Step 3).

In our basketball player example, we have 10 observations, so $n = 10$. The R-Squared from the auxiliary regression is $R^2_{\text{new}} = 0.600395$. Applying these values to the formula yields the

following test statistic:

$$F = 10 \times 0.600395 = 6.00395$$

This calculated value of 6.00395 is the crucial metric we will use to determine the p-value and ultimately decide whether to reject the null hypothesis of homoscedasticity.

Step 5: Determining the P-value and Interpreting Results

The final, decisive step is comparing the calculated test statistic (6.00395) against the critical values of the chi-squared distribution, which is used to determine the associated p-value. The degrees of freedom (df) for this test are equal to the number of independent variables used in the original regression, which is 3 in our case (Points, Assists, Rebounds).

We use Excel's built-in function, `CHISQ.DIST.RT`, which calculates the right-tailed probability (p-value) for the chi-squared distribution:

$$=CHISQ.DIST.RT(\text{test statistic}, \text{degrees of freedom})$$

Substituting our calculated values, the Excel formula is:

$$=CHISQ.DIST.RT(6.00395, 3) = \mathbf{0.111418}$$

The resulting p-value is approximately 0.111418.

Conclusion: We compare the calculated p-value (0.111418) to the conventional significance level ($\alpha = 0.05$). Since the p-value is greater than the significance level ($0.111418 > 0.05$), we **fail to reject the null hypothesis**. This means that we do not have sufficient statistical evidence at the 5% significance level to conclude that heteroscedasticity is present in the original multiple linear regression model. The variance of the error terms in the model relating player rating to performance statistics is acceptably constant, confirming the reliability of the standard errors derived from the OLS estimates.