

# Breusch-Pagan Test in Excel: A Step-by-Step Tutorial for Detecting Heteroscedasticity

Authored by  
**Mohammed looti**

November 8, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Breusch-Pagan Test in Excel: A Step-by-Step Tutorial for Detecting Heteroscedasticity*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13524>

The [Breusch-Pagan Test](#) is one of the most fundamental diagnostic procedures utilized in [regression analysis](#). Its primary purpose is to determine whether the variance of the error terms, known as the disturbance, remains constant across all observations--a condition called homoscedasticity. Failure to meet this requirement results in [heteroscedasticity](#), a condition that seriously violates a core assumption of the **Classical Linear Regression Model**.

When heteroscedasticity is present, the model's estimated coefficients remain unbiased, but the calculation of the standard errors becomes unreliable. This distortion leads directly to flawed inference: hypothesis tests and confidence intervals derived from the model cannot be trusted, potentially causing researchers to draw incorrect conclusions about the significance of their predictors. Therefore, confirming the absence of heteroscedasticity is a critical prerequisite for robust statistical modeling.

While dedicated statistical software automates this check, performing the [Breusch-Pagan Test](#) manually within a common spreadsheet environment like **Microsoft Excel** offers invaluable insight into its underlying mechanism. This comprehensive tutorial walks you through every necessary step, from initial model fitting to the final statistical decision, ensuring you can accurately diagnose your model's stability using readily available tools.

## Example Setup: Structuring Data for Statistical Diagnosis

To demonstrate the methodology clearly, we will employ a practical, small-scale dataset consisting of ten observations related to basketball player performance. Our objective is to fit a model and then rigorously test if the relationship between the chosen variables exhibits constant error variance, confirming the data's suitability for standard statistical inference.

We will construct a [multiple linear regression](#) model where the player's **Rating** serves as the dependent variable (Y). The explanatory variables (X) used to predict this rating are **Points**, **Assists**, and **Rebounds**. Once this initial model is fitted, the subsequent Breusch-Pagan procedure will focus entirely on the residuals generated by this model, assessing whether their variance is systematically related to the predictors, which would signify the presence of [heteroscedasticity](#).

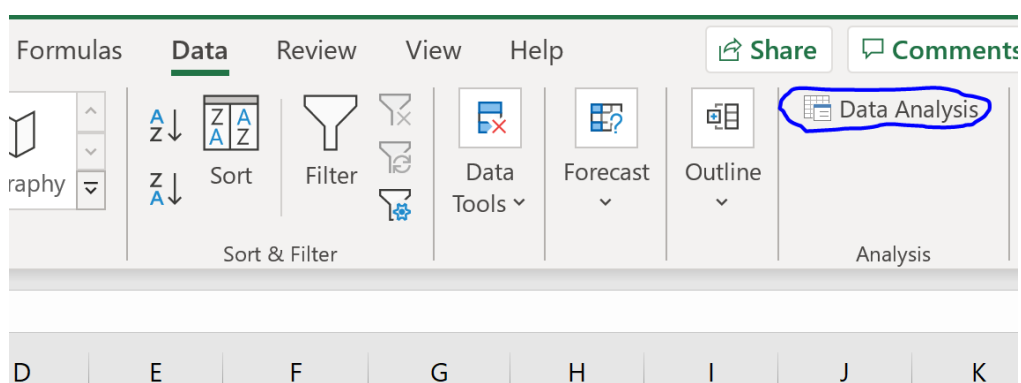
The raw data, organized and ready for the initial regression analysis in Excel, is displayed below. This dataset forms the foundation upon which all subsequent calculations--predicted values, squared residuals, and the auxiliary regression--will be built.

	A	B	C	D	E	F
1	<b>rating</b>	<b>points</b>	<b>assists</b>	<b>rebounds</b>		
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 prerequisite for the Breusch-Pagan Test is fitting the primary regression model to obtain the coefficients and, crucially, the initial set of [residuals](#). In Excel, this is most efficiently accomplished using the **Data Analysis Toolpak**. If you cannot locate the Data Analysis option under the Data tab, you must first enable this powerful statistical add-in through Excel's options menu.

To initiate the regression process, navigate to the **Data** tab and select **Data Analysis**. If the option is currently unavailable, ensure the necessary Add-In is activated before proceeding:



In the Data Analysis dialog box, select the *Regression* option. Carefully define the **Input Y Range** (our dependent variable, Rating) and the **Input X Range** (our independent variables: Points, Assists, and Rebounds). While the Toolpak can generate residuals directly, we will proceed to manually calculate the predicted values and squared residuals in the next step to ensure absolute precision, using the coefficients provided in the summary output.

	A	B	C	D	E	F	G
1	<b>rating</b>	<b>points</b>	<b>assists</b>	<b>rebounds</b>			
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

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

The output sheet provides the critical coefficients--the Intercept and the slopes for each predictor--which define the model's line of best fit. These coefficient values are non-negotiable inputs for the subsequent calculation of the predicted values, which, in turn, are essential for determining the error terms that drive the entire Breusch-Pagan methodology.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	rating	points	assists	rebounds										
2	90	25	5	11		SUMMARY OUTPUT								
3	85	20	7	8										
4	82	14	7	10		<i>Regression Statistics</i>								
5	88	16	8	6		Multiple R	0.789024							
6	94	27	5	6		R Square	0.622559							
7	90	20	7	9		Adjusted R	0.433839							
8	76	12	6	6		Standard E	4.584449							
9	75	15	9	10		Observatio	10							
10	87	14	9	10										
11	86	19	5	7		ANOVA								
12							<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
13						Regression	3	207.997	69.33232228	3.298841616	0.099468483			
14						Residual	6	126.103	21.01717219					
15						Total	9	334.1						
16														
17							<i>Coefficients</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>
18						Intercept	62.47163	14.58822	4.28233534	0.005192295	26.77555401	98.16771	26.77555	98.16771046
19						X Variable	1.119326	0.410882	2.724201924	0.03445006	0.113933593	2.124719	0.113934	2.124719302
20						X Variable	0.883401	1.380667	0.639836482	0.545919049	-2.494969629	4.261772	-2.49497	4.261772073
21						X Variable	-0.42777	0.851009	-0.502663446	0.633113984	-2.510116271	1.654574	-2.51012	1.654573647
22														

## Step 2: Deriving Predicted Values and Squaring the Residuals

The next preparatory phase involves constructing the response variable for the upcoming auxiliary regression. This response variable is the set of squared [residuals](#), representing the estimated error variance for each observation. This step requires two distinct calculations: determining the predicted value (Y?) and then calculating the squared error term.

First, we calculate the predicted value (Y?) for each observation using the standard regression equation and the coefficients derived in Step 1. This formula estimates the player's Rating based on their specific values for Points, Assists, and Rebounds. It is vital to use absolute cell references (using the dollar sign, e.g., \$A\$1) for the coefficient values to ensure the formula can be correctly copied down the column:

	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															

Applying this formula across all data points yields a column of predicted ratings. This prediction is crucial because the raw residual is defined precisely as the deviation between the observed actual rating (Y) and this calculated predicted rating (Y - Y?).

	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															

The final sub-step is calculating the squared [residuals](#). We square the raw residuals for two main reasons: mathematically, it ensures all error terms are positive, and statistically, it emphasizes larger errors, which are the primary evidence used to detect non-constant variance. The calculation

is straightforward:  $(\text{Actual Rating} - \text{Predicted Rating})^2$ . This newly calculated variable will serve as the dependent variable in the forthcoming auxiliary regression, allowing us to model the variance itself.

	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																

This column of squared residuals effectively quantifies the magnitude of the error variance associated with each observation. By using this as the response variable, we are preparing to test if this variance systematically changes as a function of the original explanatory variables.

	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																

### Step 3: Executing the Auxiliary Regression

The core of the Breusch-Pagan methodology lies in the auxiliary regression. This secondary regression tests a simple concept: if the variance of the error terms (represented by the squared residuals) can be statistically predicted by the original independent variables, then the variance is not constant, confirming [heteroscedasticity](#). Conversely, if the explanatory variables have no predictive power over the squared residuals, the variance is constant (homoscedastic).

We use the Data Analysis Toolpak for regression once more, but with a crucial modification to the input ranges:

The **Input Y Range** is now the column containing the **Squared Residuals** (the calculated error variance).

The **Input X Range** remains the original set of **explanatory variables** (Points, Assists, and Rebounds).

The specific output statistic we need from this auxiliary regression is the **Coefficient of Determination**, commonly referred to as [R-squared](#) ( $R^2_{new}$ ). A high [R-squared](#) here suggests that the original predictors successfully explain a significant portion of the variability in the error terms, indicating non-constant variance.

The summary output for this auxiliary regression, where squared residuals predict the predictors, is displayed below:

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 critical summary, we extract the [R-squared](#) value, which is approximately 0.600395. This value,  $R^2_{new}$ , is the single required input for the final calculation of the [Breusch-Pagan Test](#)

statistic.

#### Step 4: Computing the Test Statistic and P-Value

The final step involves transitioning the auxiliary regression result into a formal statistical test statistic, which follows a [Chi-Square test statistic](#) distribution. This comparison is necessary to determine the p-value and make a formal decision regarding the [null hypothesis](#) of homoscedasticity.

The Breusch-Pagan test statistic ( $X^2$ ) is calculated by taking the product of the number of observations ( $n$ ) and the R-squared value ( $R^2_{new}$ ) obtained from the auxiliary regression:

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

In our current example, we have 10 observations ( $n = 10$ ) and an  $R^2_{new}$  of 0.600395. The resulting calculation is:

$$X^2 = 10 * 0.600395 = \mathbf{6.00395}$$

Once the test statistic is determined, we must find the corresponding p-value. The [degrees of freedom](#) (df) for this test are equal to the number of independent variables ( $k$ ) used in the original regression, which is 3 (Points, Assists, Rebounds). We use Excel's `CHISQ.DIST.RT` function, which computes the right-tail probability of the Chi-Square distribution, to find the p-value:

The formula applied in Excel is:

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

#### Conclusion: Interpreting the Breusch-Pagan Test Results

The final step in the diagnostic process is to compare the calculated p-value against a predetermined significance level ( $\alpha$ ), typically 0.05. The test is framed by the following competing hypotheses:

**Null Hypothesis (H<sub>0</sub>):** The variance of the errors is constant (Homoscedasticity).

**Alternative Hypothesis (H<sub>1</sub>):** The variance of the errors is not constant (Presence of [Heteroscedasticity](#)).

Our analysis yielded a p-value of 0.111418. Since this calculated p-value (0.111418) is greater than our chosen significance threshold of 0.05, we **fail to reject the null hypothesis**. This statistical outcome indicates that there is insufficient evidence, at the 5% significance level, to conclude that heteroscedasticity is present in the [multiple linear regression](#) model. Consequently, the assumption of constant error variance is satisfied, confirming that the standard error estimates and all associated inference tests derived from the original model are statistically reliable.