How to Perform a Breusch-Pagan Test in Excel to Detect Heteroscedasticity

Authored by

stats writer

March 9, 2026

The **Breusch-Pagan test** serves as a sophisticated diagnostic tool within the field of **econometrics** and **statistics**. Its primary objective is to evaluate whether the variance of the errors in a **regression model** is constant, a condition known as **homoscedasticity**. When this assumption is violated, the model exhibits **heteroscedasticity**, which can lead to inefficient estimates and biased standard errors, potentially invalidating **hypothesis tests**.

In a typical **regression analysis**, practitioners assume that the error term--the difference between the observed and predicted values--maintains a stable variance across all levels of the independent variables. The Breusch-Pagan test formally investigates this by regressing the **squared residuals** from the original model against the independent variables. This procedure determines if the explanatory variables have a systematic relationship with the error variance, thereby identifying non-constant spread in the data.

Performing this test within **Microsoft Excel** is a highly accessible method for researchers who may not have access to specialized statistical software. By leveraging built-in functions such as the **Data Analysis Toolpak** and specific mathematical formulas, users can execute a multi-step process to derive the test statistic and its corresponding **p-value**. This ensures that the regression results are robust and that any subsequent inferences drawn from the data are statistically sound.

Perform a Breusch-Pagan Test in Excel

The **Breusch-Pagan test** is a vital diagnostic used to determine if **heteroscedasticity** is present in a **regression analysis**, which is crucial for verifying the **Gauss-Markov assumptions**.

This comprehensive tutorial provides a step-by-step guide on how to perform a Breusch-Pagan Test in Excel, ensuring your **Ordinary Least Squares** (OLS) results are reliable.

The Significance of Detecting Heteroscedasticity

Before diving into the technical execution, it is important to understand why **heteroscedasticity** matters in **data analysis**. In many real-world scenarios, the variance of the **residuals** changes as the value of an independent variable changes; for instance, income inequality often results in higher variance in spending habits among high-income earners compared to low-income earners. If this variance is not constant, the standard errors of your regression coefficients will be incorrect, leading to misleading **confidence intervals** and **t-statistics**.

The **Breusch-Pagan test** addresses this by focusing on the "Lagrange Multiplier" (LM) statistic. By examining how well the independent variables explain the **squared residuals**, the test provides a formal **statistical significance** level to help you decide whether to stick with OLS or move toward **weighted least squares** or **robust standard errors**. Excel's flexibility allows us to manually construct this test, providing deeper insight into the underlying mechanics of the **statistical model**.

By following the subsequent steps, you will learn not only how to click the right buttons in Excel but also how to interpret the **chi-square distribution** results. This foundational knowledge is essential for any **data scientist** or analyst who relies on **quantitative research** to drive decision-making. We will use a practical example involving sports analytics to demonstrate these concepts in a clear and engaging manner.

Example: Breusch-Pagan Test in Excel

For this example, we will utilize a **dataset** that details the performance attributes of 10 professional basketball players. These attributes serve as our **variables** for the study.

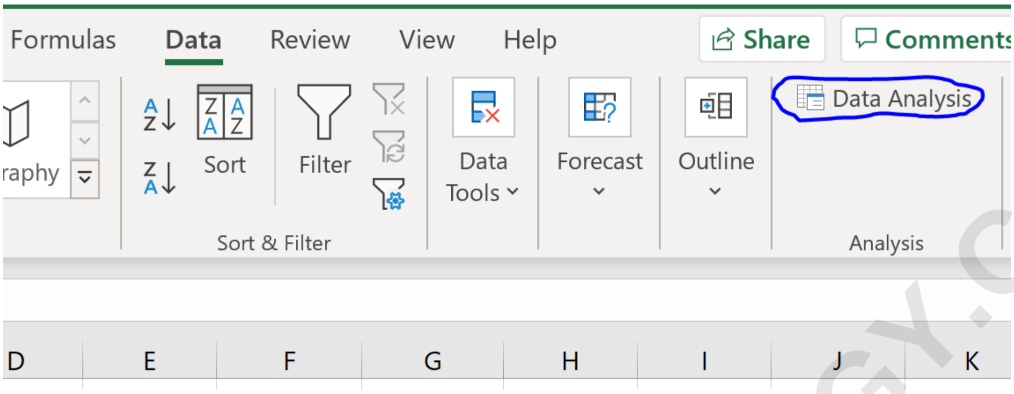
	A	B	C	D	E	F
1	rating	points	assists	rebounds		
2	90	25	5	11		
3	85	20	7	8		
4	82	14	7	10		
5	88	16	8	6		
6	94	27	5	6		
7	90	20	7	9		
8	76	12	6	6		
9	75	15	9	10		
10	87	14	9	10		
11	86	19	5	7		
12						
13						
14						
15						
16						
17						
18						

Our goal is to fit a **multiple linear regression** model. In this setup, "Rating" will serve as the **dependent variable** (response), while "Points," "Assists," and "Rebounds" will act as the **independent variables** (explanatory variables). After constructing the model, we will apply the Breusch-Pagan Test to confirm the stability of the error variance.

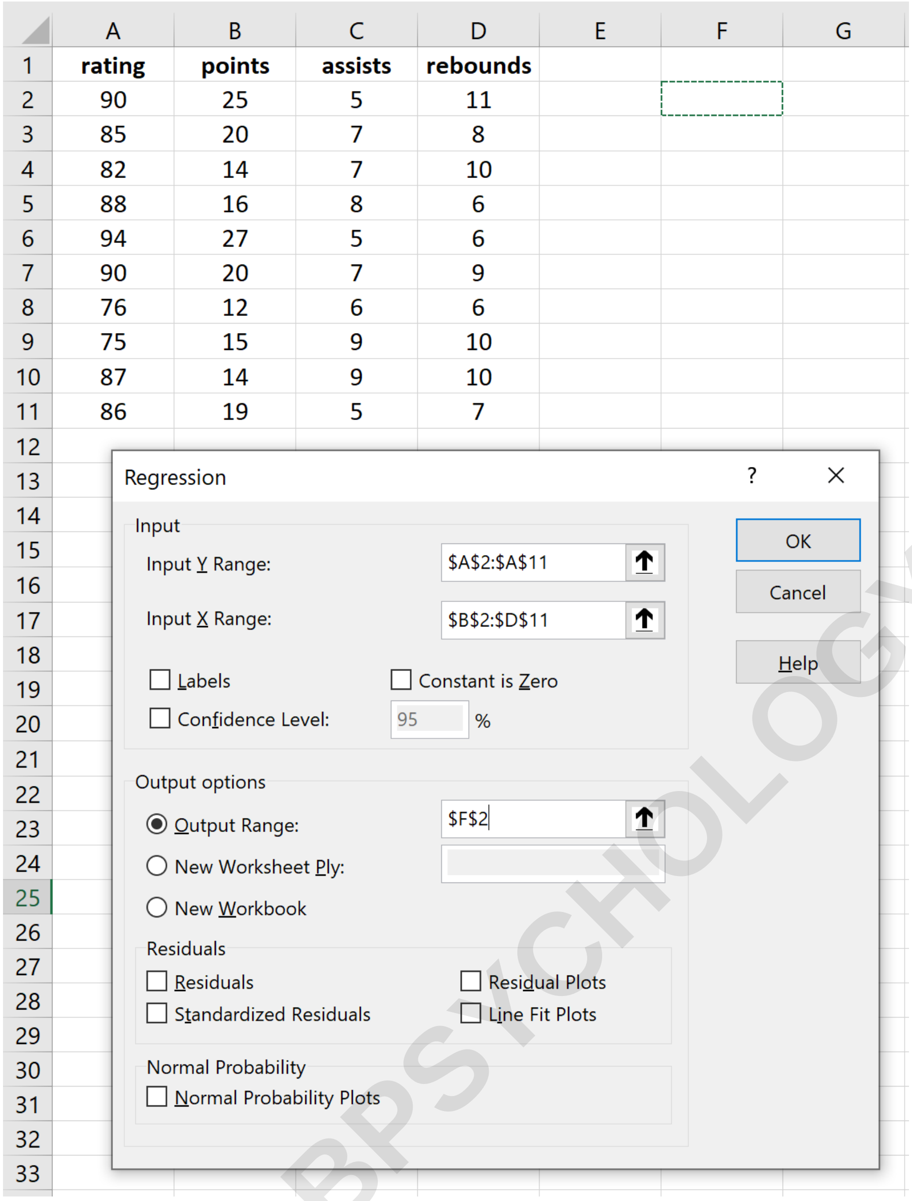
Identifying **heteroscedasticity** in this context might reveal that the model predicts player ratings more accurately for certain types of players than others. For example, if the variance of the **residuals** increases with the number of "Points" scored, it suggests that scoring is an inconsistent predictor of overall "Rating." Using the Breusch-Pagan approach allows us to quantify this inconsistency and determine if the **regression** model requires adjustment for greater accuracy.

Step 1: Perform Initial Multiple Linear Regression

To begin the analysis, navigate to the top ribbon in your Excel interface. Select the **Data** tab and locate the **Data Analysis** command. If this option is not visible, you must first enable the **Analysis Toolpak** through the Excel Options menu.



Upon clicking **Data Analysis**, a dialogue window will appear. From the list of available **statistical** tools, select ***Regression*** and click **OK**. You will then be prompted to input the "Input Y Range" (Rating) and the "Input X Range" (Points, Assists, and Rebounds). Ensure you include labels if you have selected the headers.



This process generates a comprehensive **regression output**, which includes the **R-squared** value, **ANOVA** table, and **coefficients**. These elements are the building blocks for our subsequent calculations of **residuals**.

	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		*Regression Statistics*								
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							*df*	*SS*	*MS*	*F*	*Significance F*			
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							*Coefficients*	*andard Errc*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Upper 95.0%*
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														

The output provides the **intercept** and the slopes for each explanatory variable. These **coefficients** represent the estimated relationship between each basketball metric and the player's rating. While these values tell us the direction and strength of the relationships, they do not yet tell us if the model's errors are distributed uniformly, which is why the **Breusch-Pagan test** is the next logical step.

Step 2: Calculate the Squared Residuals

The next phase of the **Breusch-Pagan test** involves determining the **residuals**--the difference between the observed "Rating" and the "Rating" predicted by our **linear model**. To compute these predicted values, we apply the **linear equation** derived from the regression **coefficients**:

	A	B	C	D	E
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	

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%*	*ower 95.0%*	*Jpper 95.0%*
Intercept	62.47163	14.5882158	4.282335	0.005192	26.77555	98.16771	26.77555	98.16771
X Variable	1.119326	0.41088233	2.724202	0.03445	0.113934	2.124719	0.113934	2.124719
X Variable	0.883401	1.38066717	0.639836	0.545919	-2.49497	4.261772	-2.49497	4.261772
X Variable	-0.42777	0.85100939	-0.50266	0.633114	-2.51012	1.654574	-2.51012	1.654574

Applying this formula across our **sample** allows us to generate a column of predicted ratings. Each predicted value represents where the player's rating "should" be according to the **mathematical model** based on their points, assists, and rebounds.

	A	B	C	D	E
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

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%*	*ower 95.0%*	*Jpper 95.0%*
Intercept	62.47163	14.5882158	4.282335	0.005192	26.77555	98.16771	26.77555	98.16771
X Variable	1.119326	0.41088233	2.724202	0.03445	0.113934	2.124719	0.113934	2.124719
X Variable	0.883401	1.38066717	0.639836	0.545919	-2.49497	4.261772	-2.49497	4.261772
X Variable	-0.42777	0.85100939	-0.50266	0.633114	-2.51012	1.654574	-2.51012	1.654574

Once the predicted values are established, we calculate the **residuals** by subtracting the predicted value from the actual observed value. To focus on the magnitude of the error variance, we then square each **residual**. Squaring is essential because it removes negative signs and emphasizes larger deviations, which is a key requirement for the **Breusch-Pagan** methodology.

rating	points	assists	rebounds	predicted	squared residuals
90	25	5	11	90.16632	=(E2-A2)^2
85	20	7	8	87.6198	
82	14	7	10	80.0483	
88	16	8	6	84.88144	
94	27	5	6	94.54382	
90	20	7	9	87.19203	
76	12	6	6	78.63733	
75	15	9	10	82.93443	
87	14	9	10	81.8151	
86	19	5	7	85.16144	

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	*andard Errc*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Jpper 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

The resulting **squared residuals** effectively represent the "variance" of each individual observation's error. If these values appear to grow or shrink systematically with our independent variables, it is a strong indicator of **heteroscedasticity**.

rating	points	assists	rebounds	predicted	squared residuals
90	25	5	11	90.16632	0.027661
85	20	7	8	87.6198	6.863348
82	14	7	10	80.0483	3.809141
88	16	8	6	84.88144	9.725433
94	27	5	6	94.54382	0.295745
90	20	7	9	87.19203	7.884707
76	12	6	6	78.63733	6.955505
75	15	9	10	82.93443	62.955129
87	14	9	10	81.8151	26.883184
86	19	5	7	85.16144	0.703180

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	*andard Errc*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Jpper 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

By organizing these **data structures** clearly in Excel, you create a transparent workflow that is easy to audit. This level of detail is particularly useful when presenting findings to stakeholders who may want to see the "raw" **error analysis** behind the final test results. The **squared residuals** will now act as the **dependent variable** in our next step.

Step 3: Performing the Auxiliary Regression

The core of the **Breusch-Pagan test** relies on what is known as an "auxiliary regression." In this step, we perform a second **multiple linear regression**. However, instead of using "Rating" as our response, we use the **squared residuals** we just calculated. The explanatory variables--Points, Assists, and Rebounds--remain exactly the same as in the first model.

This secondary **regression** tests the **null hypothesis** that the error variance is independent of the explanatory variables. If the independent variables have significant explanatory power over the **squared residuals**, the **R-squared** value of this new model will be high, suggesting the presence of **heteroscedasticity**.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.774852
R Square	0.600395
Adjusted R Square	0.400593
Standard Error	14.96632
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regression	3	2019.241	673.0802155	3.004945	0.116639
Residual	6	1343.945	223.9908692		
Total	9	3363.186			

	Coefficient	Standard E	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0	Upper 95.0%
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

In the output above, pay close attention to the **R-squared** value (often denoted as R^2). This value represents the proportion of variance in the **squared residuals** that can be explained by our basketball metrics. A higher R^2 in this auxiliary regression indicates that the "errors" are not random but are instead tied to the values of our independent variables, which is the definition of **heteroscedasticity**.

This approach is mathematically elegant because it transforms a complex problem about error distribution into a familiar **linear regression** problem. By using Excel's **Data Analysis Toolpak** for a second time, you are effectively letting the software calculate the **Lagrange Multiplier** components for you. This makes the Breusch-Pagan test highly intuitive and reproducible for any **correlation** study.

Step 4: Executing the Breusch-Pagan Test Statistic Calculation

With the auxiliary regression complete, we can now calculate the formal **test statistic** for the **Breusch-Pagan test**. This is typically done using the LM (Lagrange Multiplier) formula, which is calculated as the product of the **sample size** and the **R-squared** value from the "new" regression.

The mathematical representation is as follows:

X2 = n * R2new

In this **equation**:

n represents the total number of observations in your **sample** (in our basketball example, n = 10).
R2new is the **Coefficient of Determination** from the auxiliary regression where squared residuals were the dependent variable.

Applying this to our specific dataset: **X2 = 10 * 0.600395 = 6.00395**. This value is our **chi-square test statistic**. It measures the discrepancy between our data and the assumption of **homoscedasticity**. The larger this statistic, the more likely it is that **heteroscedasticity** exists within the **regression model**.

Calculating this manually in Excel helps solidify your understanding of how the **sample size** influences the **power** of the test. A larger sample size will make the test more sensitive to even small amounts of non-constant variance. Conversely, with a small sample size like ours (n=10), the test requires a much stronger relationship between the variables and the **residuals** to reach **statistical significance**.

Step 5: Determining the P-Value and Interpreting the Results

The final step in the **Breusch-Pagan test** is to determine the **p-value** associated with our calculated **test statistic**. The p-value tells us the **probability** of observing such a result if the **null hypothesis** (that homoscedasticity exists) were actually true.

To find this in Excel, we use the **chi-square distribution** function for the right tail:

=CHISQ.DIST.RT(test_statistic, degrees_of_freedom)

The **degrees of freedom** (df) for this test is equal to the number of independent variables in your model. In our example, we have three explanatory variables (Points, Assists, and Rebounds), so **df = 3**. Entering the values into the **function** yields:

=CHISQ.DIST.RT(6.00395, 3) = 0.111418

To interpret this result, we compare the **p-value** to a predetermined **significance level** (usually alpha = 0.05). Since **0.111418** is greater than **0.05**, we **fail to reject the null hypothesis**. This means we do not have sufficient evidence to conclude that **heteroscedasticity** is present in our basketball **regression** model.

While the test indicates that **homoscedasticity** is a reasonable assumption for this specific **dataset**, researchers should always remain cautious. A non-significant p-value does not "prove" that the variance is perfectly constant; it simply means that the evidence of **heteroscedasticity** is not strong enough to be **statistically significant** at the chosen level. This distinction is vital for maintaining **scientific rigor** in any **quantitative analysis**.

Conclusion and Practical Recommendations

Successfully performing a **Breusch-Pagan test** in Excel provides a robust framework for validating the assumptions of your **linear regression**. By meticulously organizing your data, calculating **residuals**, and applying the **chi-square distribution**, you ensure that your **statistical inferences** are built on a solid foundation. This process is an essential part of the **econometrics** workflow, helping to avoid common pitfalls in **data modeling**.

If you find that your test results *do* indicate the presence of **heteroscedasticity** (i.e., the p-value is less than 0.05), there are several remedial actions you can take. One common approach is to transform the **dependent variable** using a **logarithmic transformation**, which often stabilizes the variance. Alternatively, you might employ **weighted least squares** (WLS), where observations with smaller error variances are given more weight in the **regression** than those with larger variances.

In summary, the **Breusch-Pagan test** is not just a mathematical hurdle but a diagnostic gateway to more accurate and **reliable** results. Whether you are analyzing sports data, financial trends, or social science **variables**, mastering this test in **Excel** empowers you to produce professional-grade **statistical** reports. By consistently checking for **heteroscedasticity**, you protect your research from the errors that frequently undermine **quantitative research**.