**HOMO and HETRO**

Homoscedasticity is the assumption in a regression model that the variance of the error term is constant, while heteroscedasticity means this variance is not constant. The purpose of homoscedasticity is to ensure the model's predictions are consistently reliable and the statistical inference is valid; its absence (heteroscedasticity) indicates the model's error is uneven, leading to biased estimates, incorrect standard errors, and unreliable conclusions.

Homoscedasticity

* **Purpose**: To ensure the model's error variance is constant across all values of the independent variables.
* **What it indicates**: This is a desirable state where the model is well-defined and makes predictions with consistent accuracy.
* **Significance**: It is a key assumption of many statistical models, like OLS regression, which ensures the validity of the results, especially for hypothesis testing.

Heteroscedasticity

* **Purpose**: To describe the situation where the error variance is not constant.
* **What it indicates**: The spread or dispersion of the data points is not consistent, meaning the model's predictions are less reliable in certain areas.
* **Significance**: It violates the assumptions of standard regression analysis, leading to several problems:
  + Biased or inefficient parameter estimates
  + Incorrect standard errors
  + Invalid hypothesis testing

**Conclusion**: Homoscedasticity is a desirable assumption in a regression model because it ensures constant error variance, leading to more accurate estimates and valid statistical inferences. In contrast, heteroscedasticity involves changing error variance, which can weaken the effectiveness and reliability of the model.

**CONCLUSION for the Given Dataset:**

**#Statistical Test – Breusch-Pagan Test**

from statsmodels.stats.diagnostic import het\_breuschpagan

bp\_test = het\_breuschpagan(residuals, X\_const)

labels = ['Lagrange multiplier', 'p-value', 'f-value', 'f p-value']

result = dict(zip(labels, bp\_test))

print(result)

{'Lagrange multiplier': 4.585366270520553, 'p-value': 0.4685435989941389, 'f-value': 0.9109077002418876, 'f p-value': 0.47484047550183595}

**Key Value to Focus On:**

You should look at either the p-value or f p-value - both indicate the same conclusion.

p-value is 0.4685

f p-value is 0.4748

**Hypotheses of the Breusch–Pagan Test**

* Null Hypothesis (H₀): Errors are homoscedastic (constant variance)
* Alternative Hypothesis (H₁): Errors are heteroscedastic (non-constant variance)

**Decision Rule**

* If p-value < 0.05, reject the null hypothesis (means heteroscedasticity is present).
* If p-value ≥ 0.05, fail to reject the null hypothesis (means variance is constant).

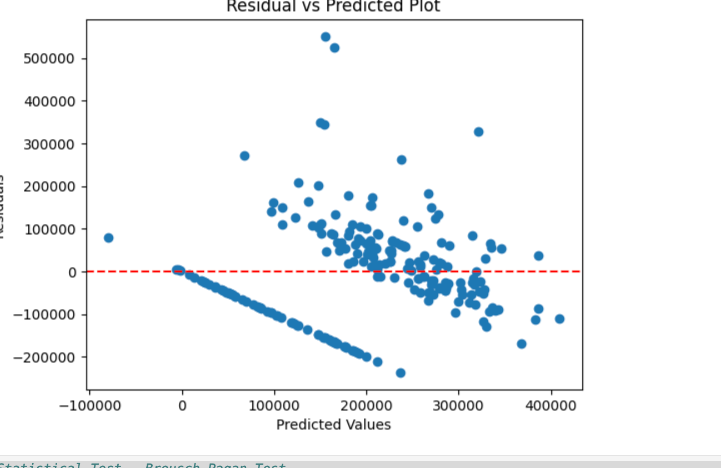
**Your Result**

* Since the p-value is 0.46 (much greater than 0.05):
* Fail to reject the null hypothesis
* No evidence of heteroscedasticity
* The model satisfies the homoscedasticity assumption

**Conclusion**:

The Breusch–Pagan test resulted in a p-value of 0.4685, which is greater than 0.05. Therefore, we fail to reject the null hypothesis and conclude that there is no significant evidence of heteroscedasticity in the residuals. The model satisfies the homoscedasticity assumption.

**Residual and Predicted Plot**



**Visible Pattern / Trend**

* The residuals are not randomly scattered.
* There seems to be a downward trend or curve, especially for middle predicted values.  
  Uneven Spread
* Residuals closer to smaller predicted values appear differently distributed from larger ones.
* Some clustering suggests that variance may not be constant.

**Possible Issues:**

* Heteroscedasticity (non-constant variance)
* Model Mis-specification
* Missing important features
* Non-linear relationships
* Outliers or influential points