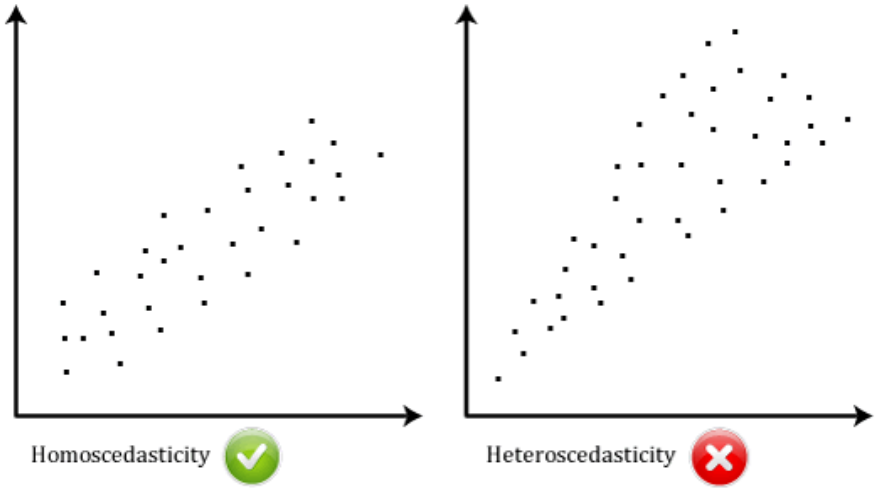
Homoscedasticity and Heteroscedasticity



Homoscedasticity:

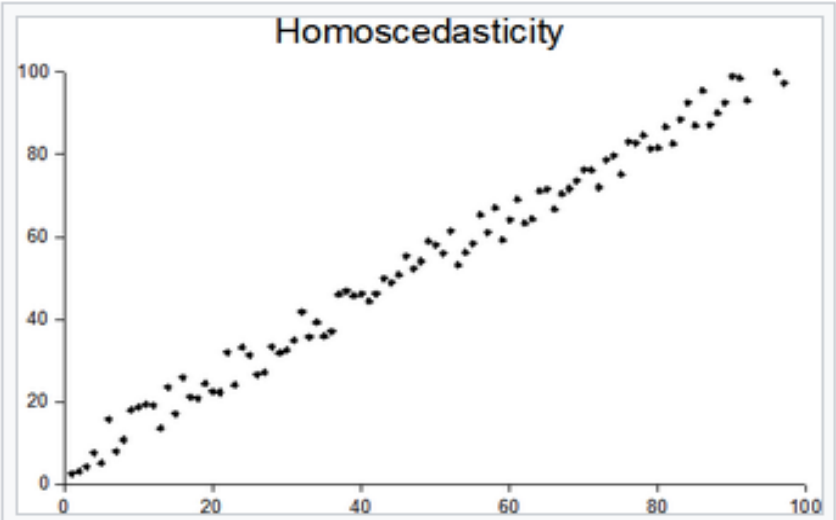
- Homoscedasticity means that the variability (or spread) of the data points around the regression line is constant across all levels of the predictor variable(s).
- In simpler terms, it implies that the scatter of data points around the regression line is consistent, forming a uniform "cloud" of points.
- When homoscedasticity is present, it indicates that the model's errors (or residuals) have constant variance, regardless of the values of the predictor variables.

When to use :

- When comparing different regression models, such as models with different sets of predictor variables or transformations, it's important to consider homoscedasticity to ensure that the models are comparable and reliable.
- If heteroscedasticity is detected, addressing it may lead to improvements in the model's performance, such as better coefficient estimates and more accurate predictions.

Real-time example of Homoscedasticity:

- Suppose you are analyzing the relationship between the number of hours students study per week and their exam scores.
- In a homoscedastic scenario, regardless of how many hours a student studies, the variability in exam scores around the regression line remains constant.
- For instance, if you plot the residuals (differences between observed and predicted exam scores) against the number of hours studied, you would observe a random scattering of points with roughly the same spread throughout the range of hours studied.



Heteroscedasticity:

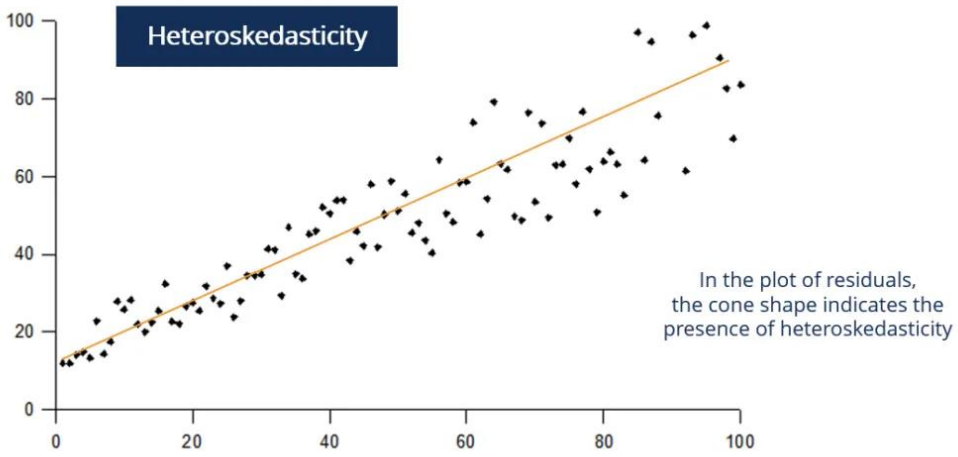
- Heteroscedasticity, on the other hand, means that the variability of the data points around the regression line is not constant across all levels of the predictor variable(s).
- In simpler terms, it suggests that the spread of data points around the regression line is not consistent, and the "cloud" of points may widen or narrow as you move along the predictor variables.
- When heteroscedasticity is present, it indicates that the model's errors (or residuals) have varying variance, which may be influenced by certain values or patterns in the predictor variables.

When to use:

- Comparing regression models, diagnosing model assumptions, evaluating prediction accuracy, or seeking to improve the reliability of your regression analysis.
- It's an essential aspect of regression analysis that ensures the validity and reliability of the results.

Real-time example of Heteroscedasticity: Continuing with the previous example, heteroscedasticity might occur if, for example, the variability in exam scores tends to increase as the number of hours studied increases. In this case, you would notice that the spread of residuals widens as you move along the x-axis, indicating that the variability of

exam scores increases with the number of hours studied.



Best One: Homo or Hetro

- Homoscedasticity (constant variance of residuals) is generally considered preferable to heteroscedasticity (varying variance of residuals).
- Heteroscedasticity is not ideal, it's important to note that in some cases, it may be unavoidable due to the nature of the data.
- In such situations, techniques such as robust standard errors or weighted least squares regression can be used to mitigate the effects of heteroscedasticity. However, whenever possible, efforts should be made to address or minimize heteroscedasticity to ensure the validity and reliability of the regression analysis.