

Understanding Homoscedasticity vs Heteroscedasticity

A beginner-friendly guide with example and rules of thumb

Homoscedasticity (✓ Good)

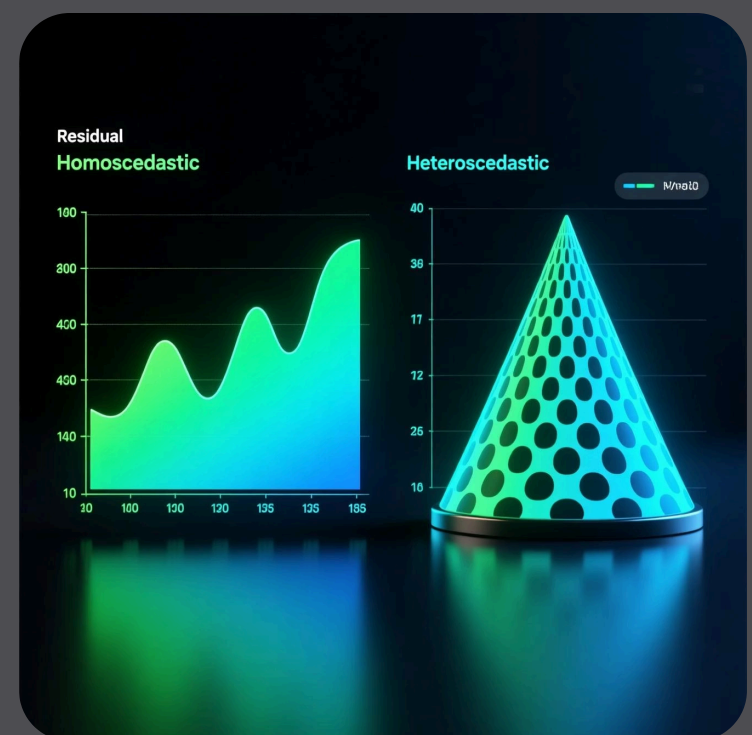
Errors (residuals) have **constant variance**.
Predictions are equally reliable for small and large values.

Heteroscedasticity (✗ Bad)

Errors (residuals) have **changing variance**.
Predictions may be stable for small values but unstable for large values.

Thumb Rule

- Residuals should look like **random noise, equally spread**.
- Cone-shaped/funnel residual plots = heteroscedasticity.
- Linear regression assumes **homoscedasticity**.



Real-time Example (House Prices)

Predicting house prices with inputs (size, rooms, location).

Homoscedastic

Prediction error is ~constant ($\pm ₹50K$) for both cheap and expensive houses.

Heteroscedastic

Cheap houses error small ($\pm ₹50K$), but luxury houses error very large ($\pm ₹15L$).

Which one is best?

✓ **Homoscedasticity is best** → regression results are reliable.

✗ **Heteroscedasticity is problematic** → standard errors & p-values unreliable.

How to fix heteroscedasticity

Apply transformations (log, sqrt).

Use Weighted Least Squares.

Apply Robust Standard Errors.