Homoscedasticity

What is Hemo?

Homoscedasticity means:

> The variance of errors (residuals) is constant across all levels of input variables. Residuals are evenly spread.

Thumb Rule (HEMO)

Residuals look like a horizontal band
Spread does NOT increase or decrease
Variance=constant

Residual plot:
Random, uniform scatter → Homoscedastic

Importance of HEMO

- Assumption of:
  - Linear Regression
  - Logistic Regression
  - ANOVA
- Gives:
  - Reliable coefficients
  - Correct p-values
  - Accurate confidence intervals

Where to Use HEMO?

Area	Reason
Medical data	Reliable prediction
Salary prediction	Fair coefficient interpretation

Area	Reason
Economics	Stable variance
Statistical inference	Valid hypothesis testing

Heteroscedasticity

What is Hetro?

Heteroscedasticity means:

> The variance of errors changes with input values.Residuals fan out or funnel in.

Thumb Rule (HETRO)

Residuals show cone / funnel / wave shape
Spread increases or decreases
Variance is not constant

Residual plot:
Patterned or widening scatter → Heteroscedastic

Why HETRO is a Problem?

Coefficients may still be unbiased

Standard errors are wrong

p-values become unreliable

Confidence intervals are misleading

Where HETRO Naturally Occurs?

Domain	Reason
Medical costs	Expenses rise with severity

Domain	Reason
Salary data	Higher salaries have more variation
Finance	Volatility clustering
Housing prices	Expensive houses vary more

Is HETRO Always Bad?

For statistical inference → BAD
For pure prediction → Acceptable

Tree models don’t care about heteroscedasticity.

How to Handle HETRO?

Method	When to Use
Log transformation	Salary, income
Weighted regression	Known variance
Robust standard errors	Statistical models
Tree-based models	Prediction tasks

HEMO vs HETRO (FINAL COMPARISON)

Feature	HEMO	HETRO
Variance	Constant	Changing
Regression assumption	Yes	Violated
Inference accuracy	High	Low
Prediction accuracy	Good	Model-dependent
Preferred	YES	Depends

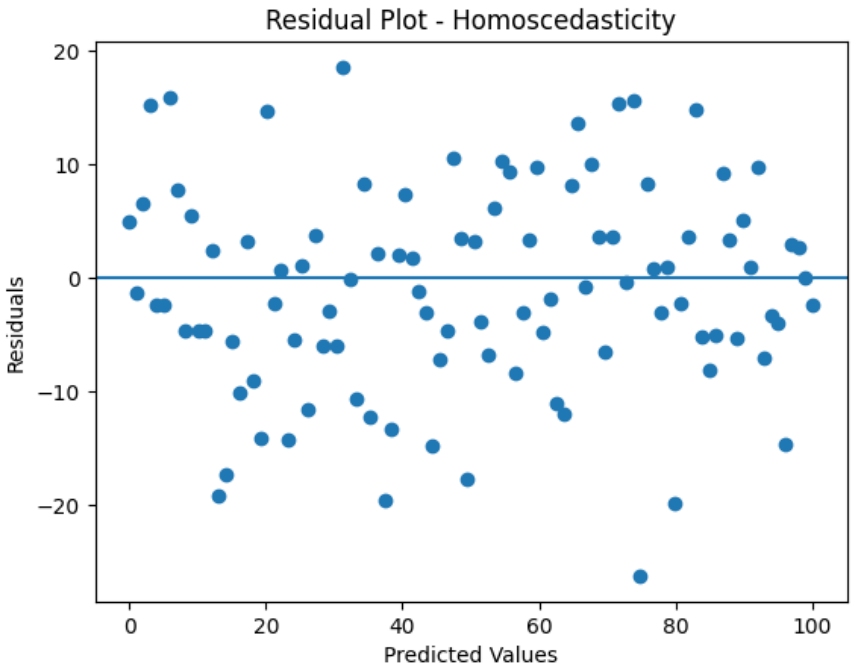
Heteroscedasticity occurs when error variance changes, affecting standard errors and hypothesis testing.

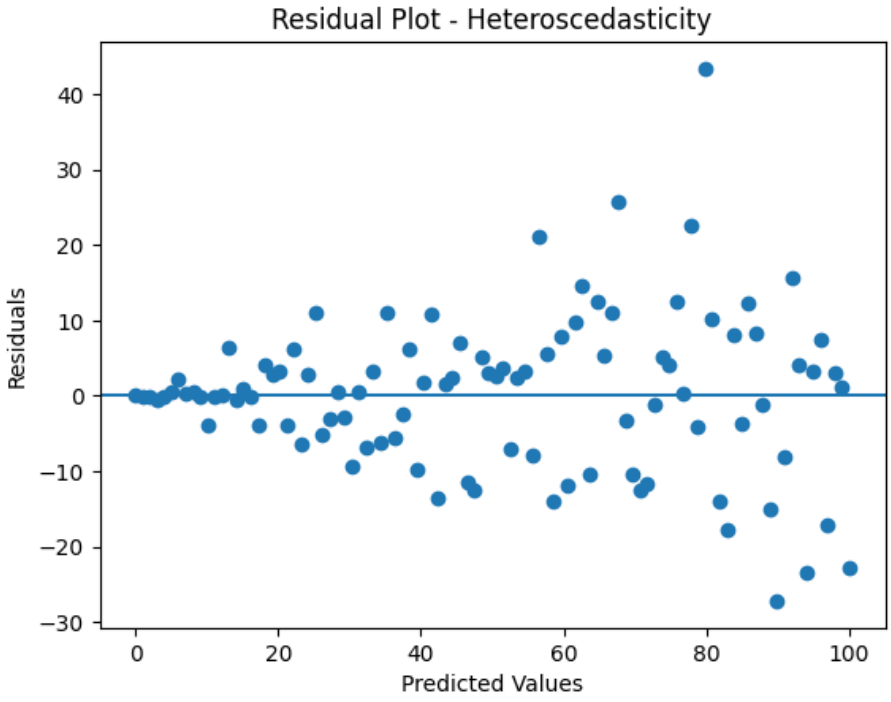
WHICH IS BETTER?

> “Homoscedasticity is preferred. Heteroscedasticity is acceptable for prediction but not for statistical inference.”

Thumb Rule to Remember Forever

If variance is constant → HEMO → GOOD
If variance changes → HETRO → PROBLEM





```
from sklearn.metrics import mean_absolute_error,
mean_squared_error
from sklearn.linear_model import LinearRegression

# Train a model (example)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

# Calculate errors
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = mean_squared_error(y_test, y_pred, squared=False) # Use
squared=False for RMSE


print(f"Mean Absolute Error: {mae}")
print(f"Mean Squared Error: {mse}")
print(f"Root Mean Squared Error: {rmse}")
```