Model Assumptions and Diagnostics

Key Assumptions

1. Linearity
True relationship is linear (or approximately)

2. Independence
Observations are independent

3. Homoscedasticity
Constant error variance across X

4. Normality
Errors ~ $N(0, \sigma^2)$

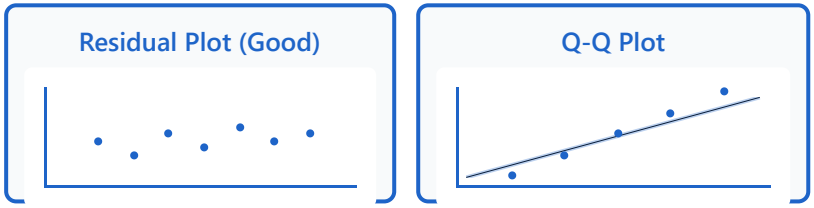
5. No Multicollinearity
Predictors not highly correlated

Diagnostic Tools

- Residual Plots
- Q-Q Plots
- Leverage Plots
- VIF Scores

Visual Diagnostics

Example Diagnostic Plots



Random scatter in residual plot = Good fit
Points on diagonal in Q-Q = Normal errors

Violations & Remedies

- Transformations (log, sqrt)
- Robust regression methods
- Regularization (Ridge, Lasso)

Model Validation

- Train-Test Split
- Cross-Validation
- Bootstrap