

Jim's Cheat Sheet

for Detecting and Correcting Linear Regression Problems



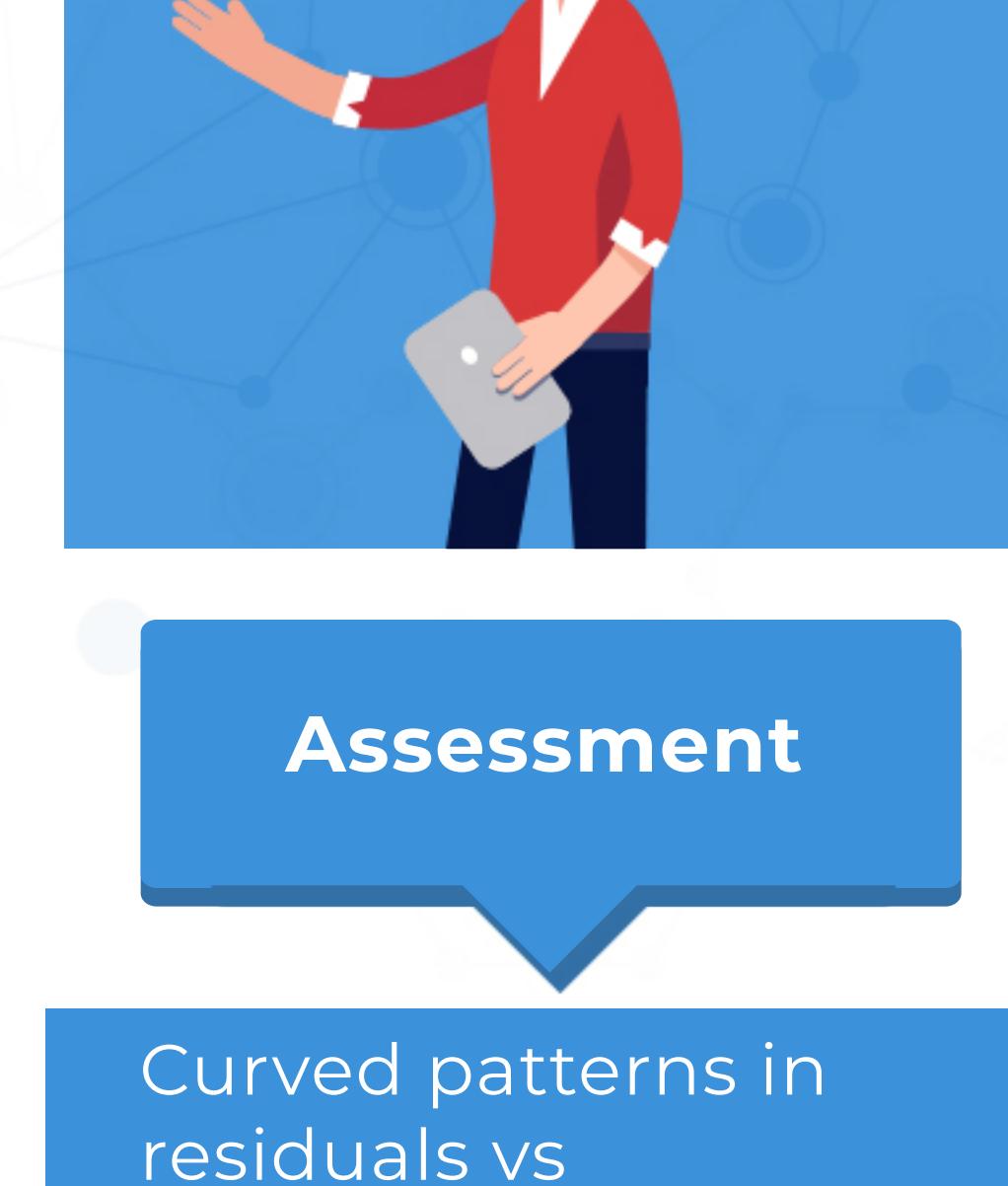
Linear regression (OLS) has a minefield of potential assumption violations and other issues that can produce results you can't trust. This cheat sheet helps you navigate the numerous problems, detection methods, and corrective measures! Be confident in your results!



OLS Potential Problem

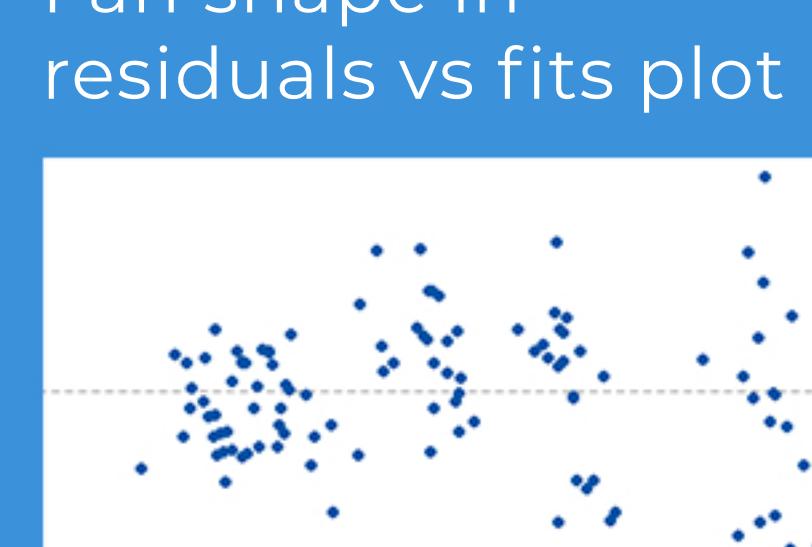
Model does not adequately fit curvature in data.

Learn about residual plots

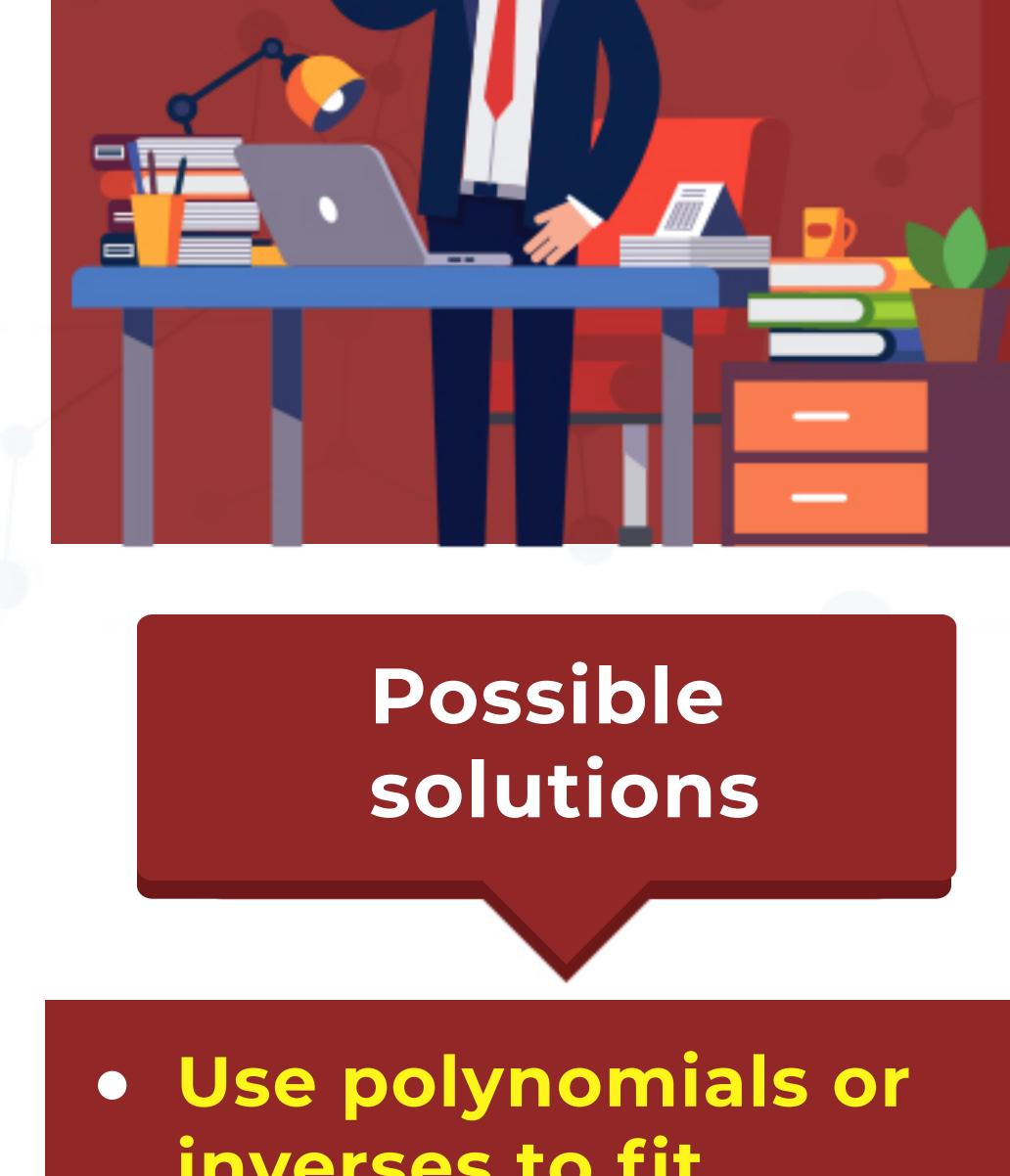


Assessment

Curved patterns in residuals vs variables plot



Learn about the OLS assumptions



Possible solutions

- Use polynomials or inverses to fit curves using OLS

- Use Nonlinear regression to fit a wider variety of curves

- Data transformation (e.g., log)

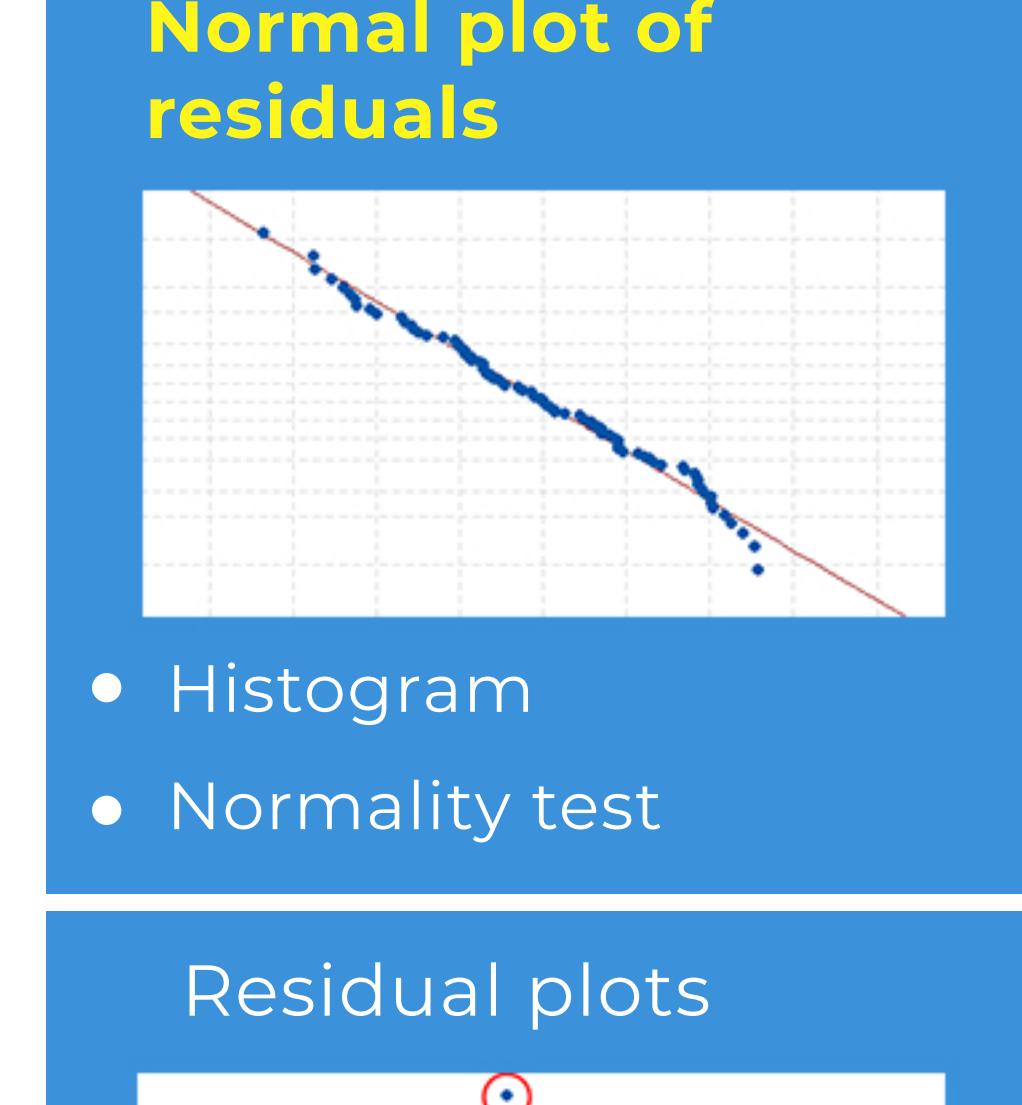
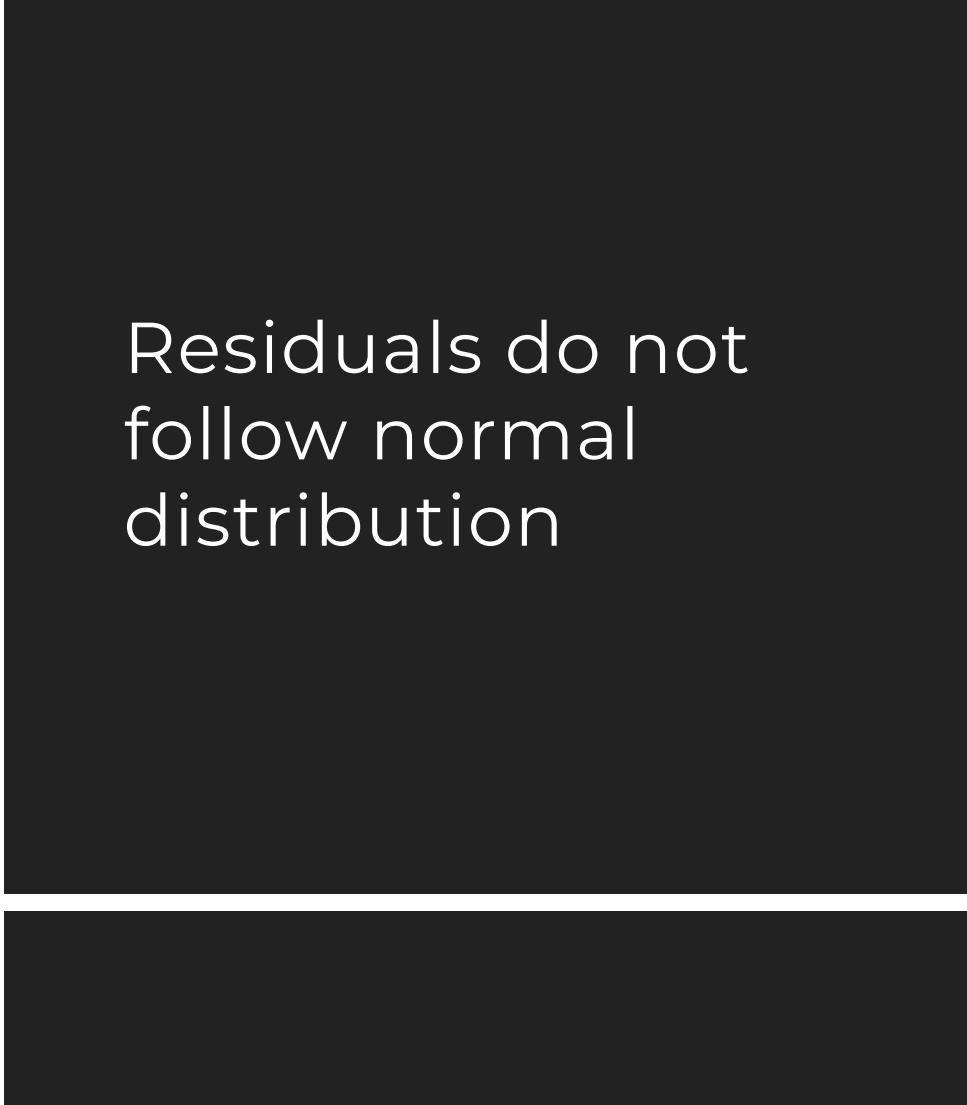
- Redefine IVs to focus on rates, per capita, etc. rather than raw measure
- Weighted least squares
- Data transformation

- Add lag variable, which are past values of the IVs
- Add independent variable to include time information
- Time series analysis

- Specify correct model
- Use Generalized Linear Models (GLM) regression which allows for other error distributions
- Data transformation and then use OLS

- Remove unusual observations when they don't reflect your study area.
- Or, include model with and without observations in report and discuss.

- Remove independent variable
- Linear combination of variables
- Use PLS, LASSO, or Ridge regression



- Fit model with and without observations to observe how it affects model estimates
- DFITS
- Cook's distance
- Leverages

- Remove unusual observations when they don't reflect your study area.
- Or, include model with and without observations in report and discuss.

- Remove independent variable
- Linear combination of variables
- Use PLS, LASSO, or Ridge regression



Visit StatisticsByJim.com where I help you intuitively understand statistics by focusing on concepts using plain English so you can concentrate on understanding your results.