conditions for linear regression

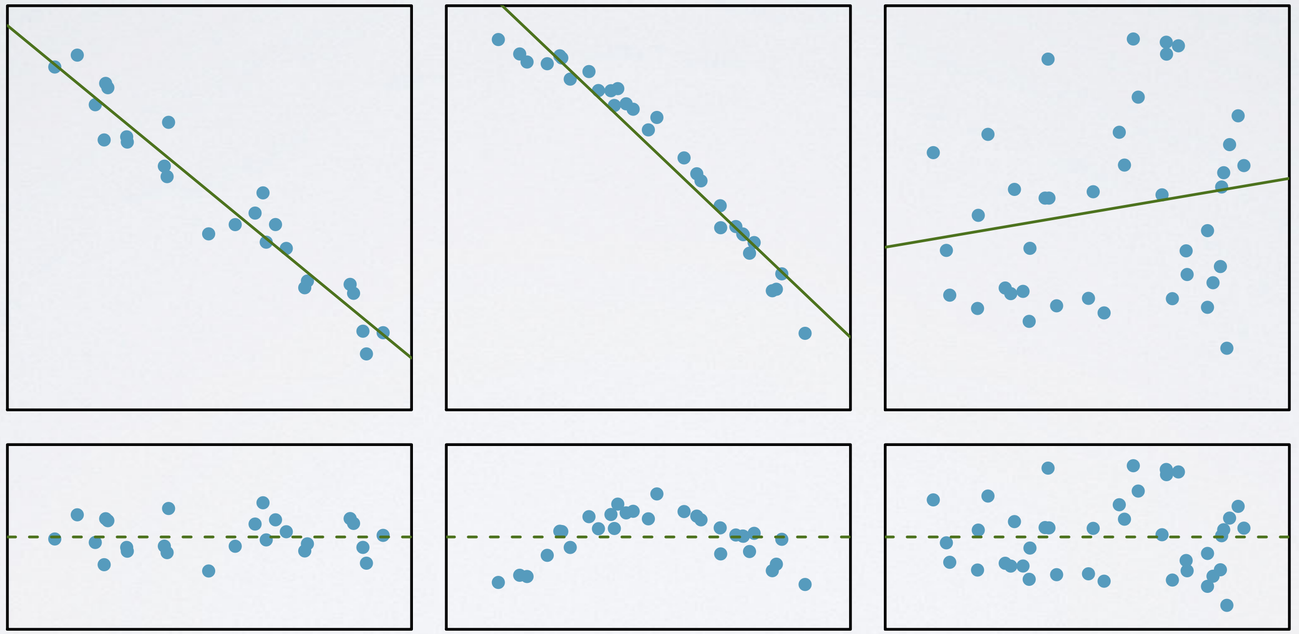
1. linearity
2. nearly normal residuals
3. constant variability

Duke UNIVERSITY

Dr. Mine Çetinkaya-Rundel
Duke University

(1) linearity

- relationship between the explanatory and the response variable should be linear
- methods for fitting a model to non-linear relationships exist
- check using a scatterplot of the data, or a residuals plot



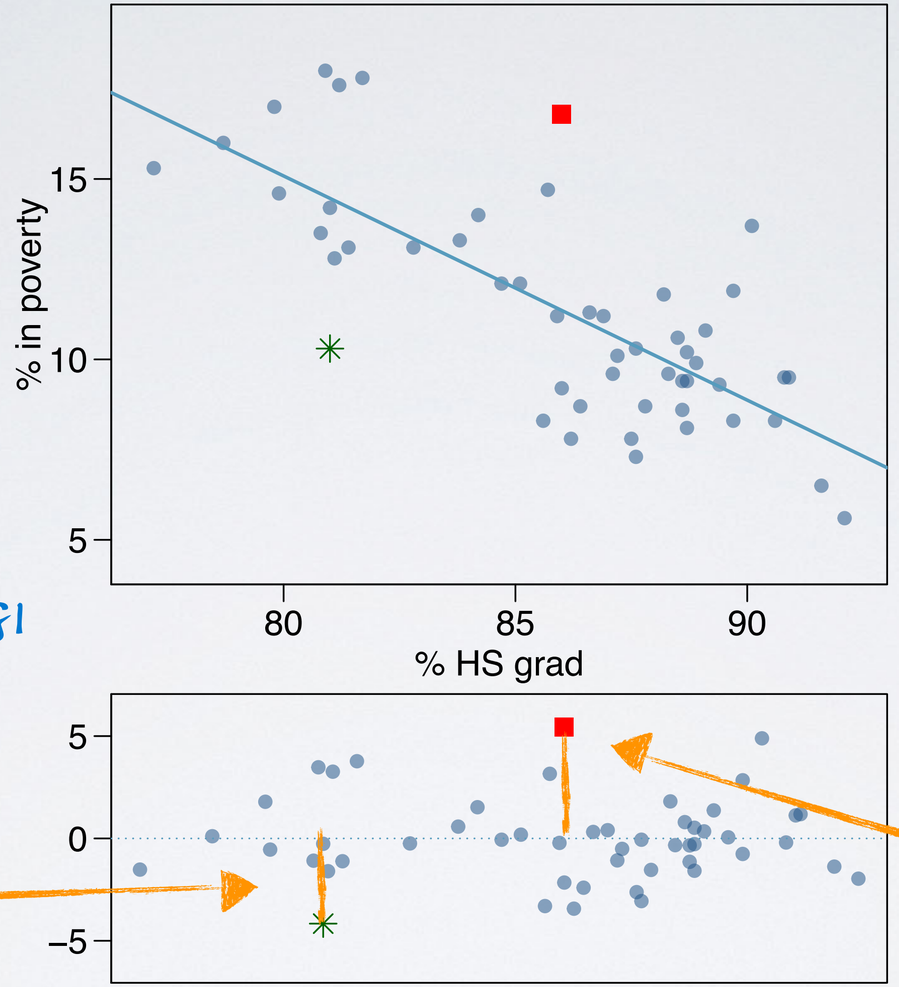
anatomy of a residuals plot

*** RI**

% HS grad = 81%
% in poverty = 10.3 %

■ DC

% HS grad = 86%
% in poverty = 16.8 %



$\widehat{\%pov} = 64.68 - 0.62 \times 81$

$= 14.46\%$

$e = 10.3 - 14.46$

$= -4.16\%$

$\widehat{\%pov} = 64.68 - 0.62 \times 86$

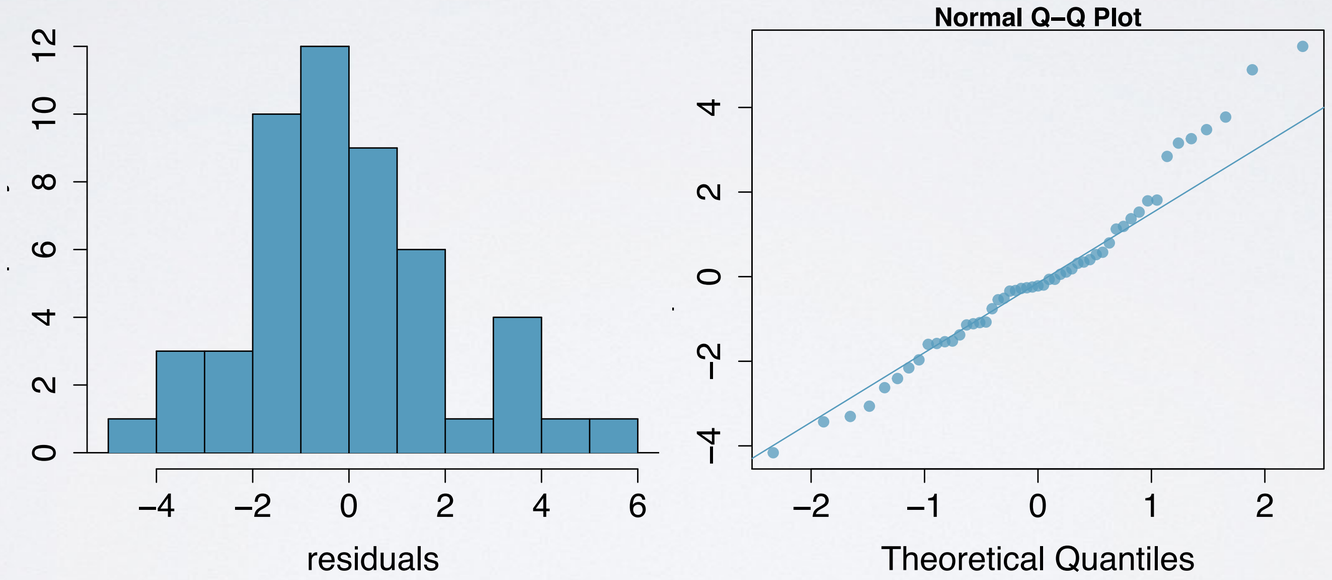
$= 11.36\%$

$e = 16.8 - 11.36$

$= 5.44\%$

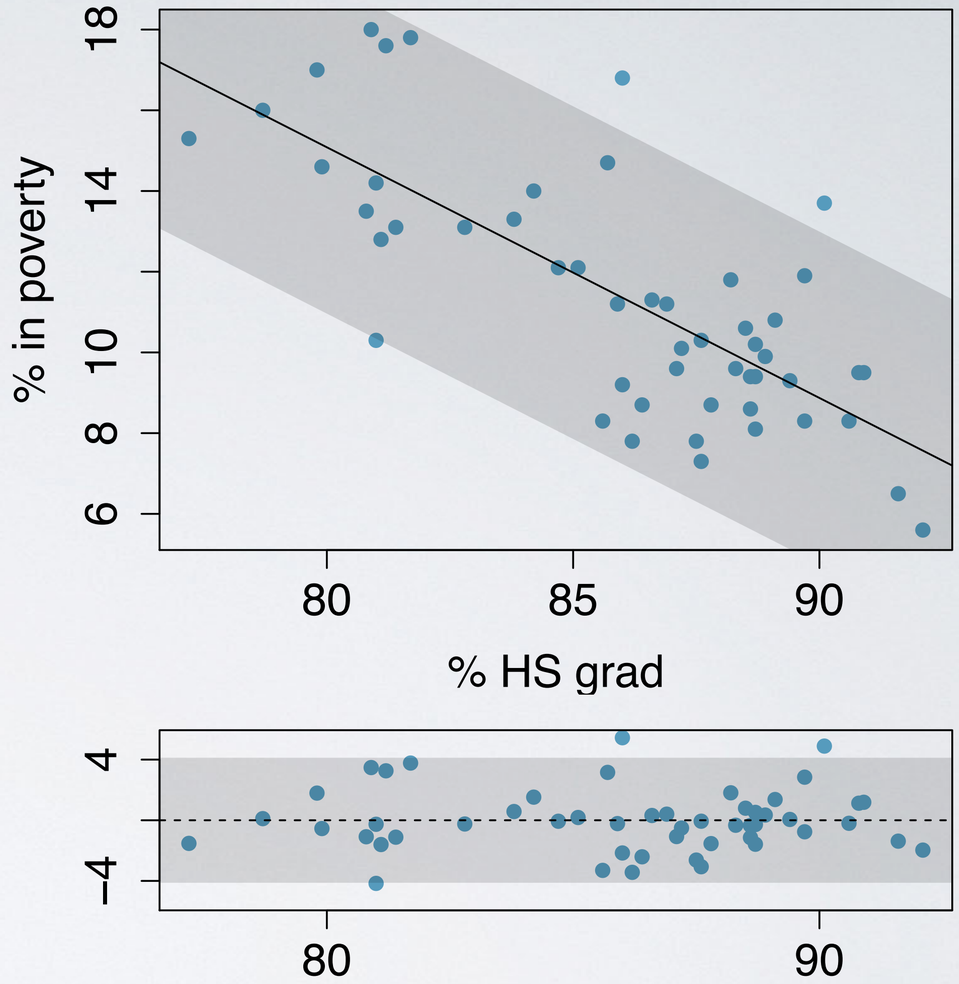
(2) nearly normal residuals

- residuals should be nearly normally distributed, centered at 0
- may not be satisfied if there are unusual observations that don't follow the trend of the rest of the data
- check using a histogram or normal probability plot of residuals



(3) constant variability

- variability of points around the least squares line should be roughly constant
- implies that the variability of residuals around the 0 line should be roughly constant as well
- also called homoscedasticity
- check using a residuals plot



Diagnostics for simple linear regression

Select a trend:

- Linear up
- Linear down
- Curved up
- Curved down
- Fan-shaped

Show residuals

This applet uses ordinary least squares (OLS) to fit a regression line to the data with the selected trend. The applet is designed to help you practice evaluating whether or not the linear model is an appropriate fit to the data. The three diagnostic plots on the lower half of the page are provided to help you identify undesirable patterns in the residuals that may arise from non-linear trends in the data.

Rate this app!

View code

Check out other apps

Want to learn more for free?

http://bitly.com/slr_diag