



DSC 10, Spring 2018

Lecture 24

Residuals

sites.google.com/eng.ucsd.edu/dsc-10-spring-2018

Residuals

(Demo)

Residuals

- Error in regression estimate
 - One residual corresponding to each point (x, y)
 - residual = observed y - regression estimate of y
 - = observed y - height of regression line at x
 - = vertical *difference* between point and line
-

Residual Plot

For **good** regressions, the residual plot

- Should look like a blob
- About half above and half below the horizontal line at 0
- Similar vertical spread throughout
- No pattern

For **bad** regressions...?

Dugong



Spotting Problems

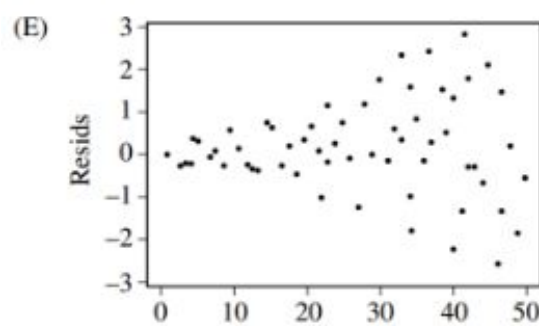
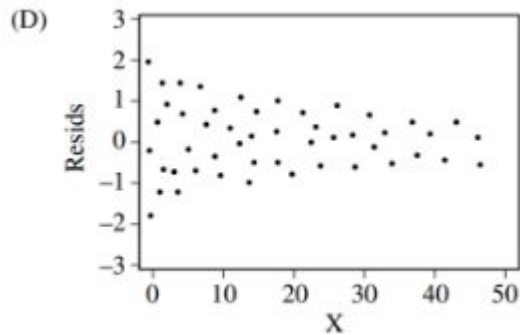
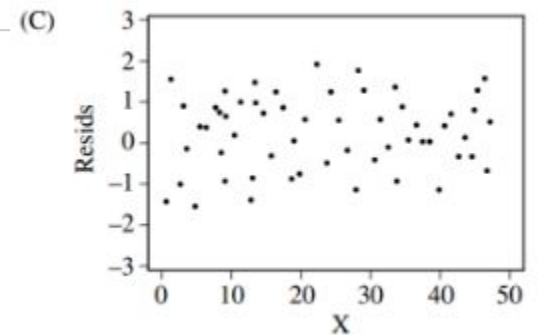
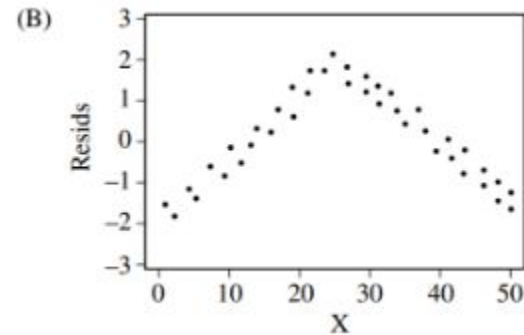
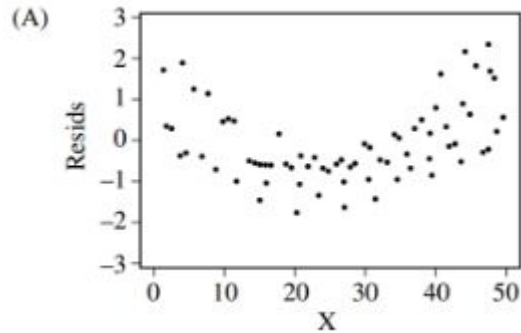
Residual plots can be used to detect:

- Non-linearity
 - Shape of scatter plot is curved, not a straight line
 - Heteroscedasticity
 - Uneven spread
-

Residual Plots are Flat Overall

- For **any** regression, no matter what the shape of the original scatterplot:
 - The residual plot will not have any linear trend, neither upwards nor downwards.
 - The correlation between the residuals and the predictor variable is 0.
-

Discussion Question

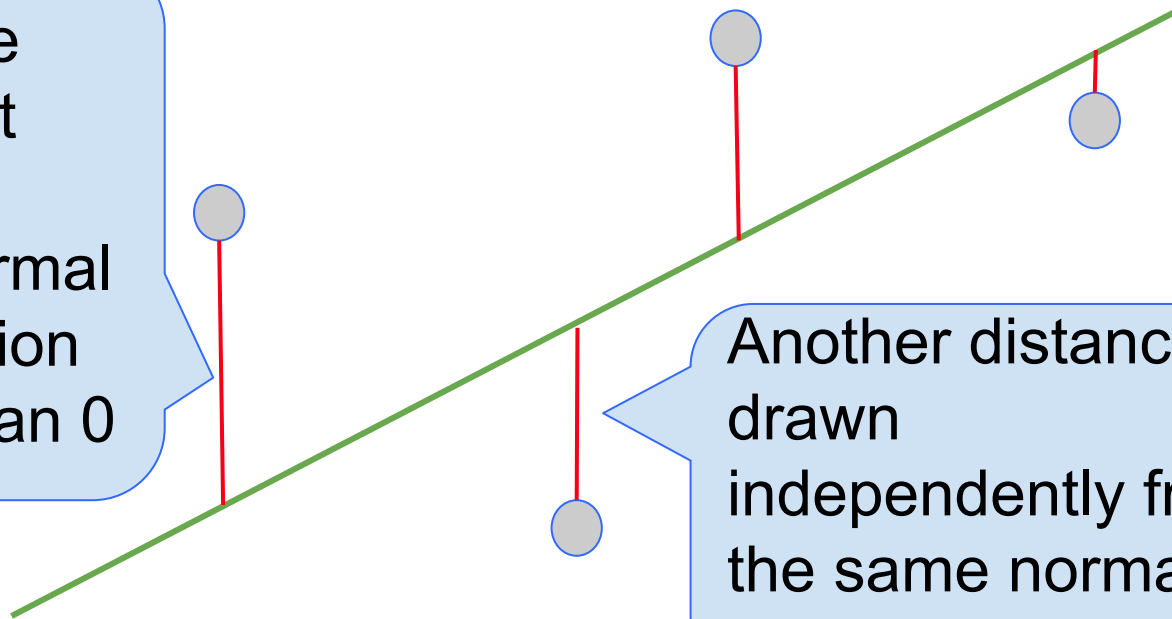


Which of the plots provides the **strongest** evidence that the regression line is an appropriate model?

Regression Model

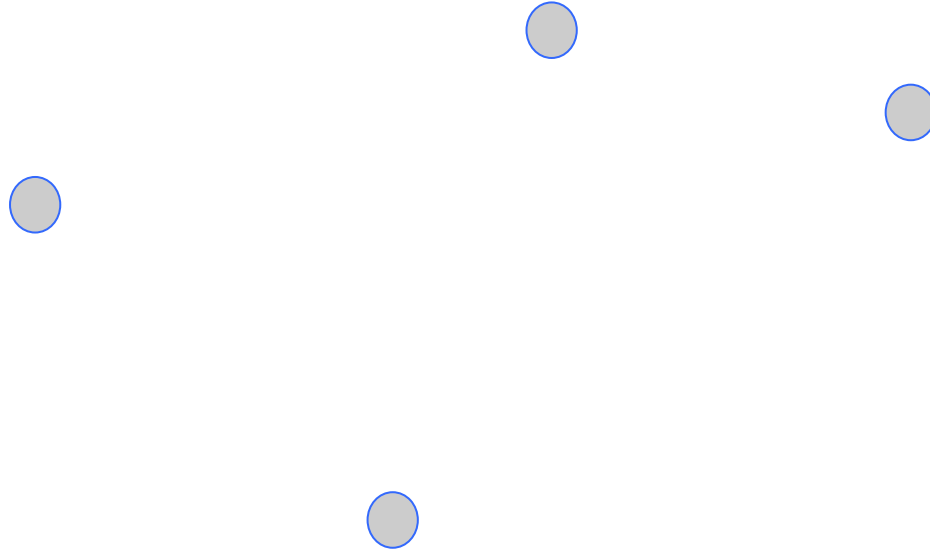
A “Model”: Signal + Noise

Distance
drawn at
random
from normal
distribution
with mean 0



Another distance
drawn
independently from
the same normal
distribution

What We Get to See



(Demo)
