190F Foundations of Data Science

Spring 2020

Lecture 19

Residuals

Announcements

Residuals

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

(Demo)

Residual Plot

A scatter diagram of residuals

- Should look like an unassociated blob for linear relations
- But still contains patterns for non-linear relations
- Used to check whether linear regression is appropriate

(Demo)

Regression Diagnostics

Dugong

Properties of Residuals

Residual Variance

The mean of residuals is always 0, regardless of the original data

Variance is standard deviation squared: mean squared deviation

(Variance of residuals) / (Variance of y) = $1 - r^2$

(Variance of fitted values) / (Variance of y) = r^2

Variance of y = (Variance of fitted values) + (Variance of residuals)

(Demo)

Discussion Question

How does the SD of the fitted values relate to r?

A. (SD of fitted) / (SD of y) = r

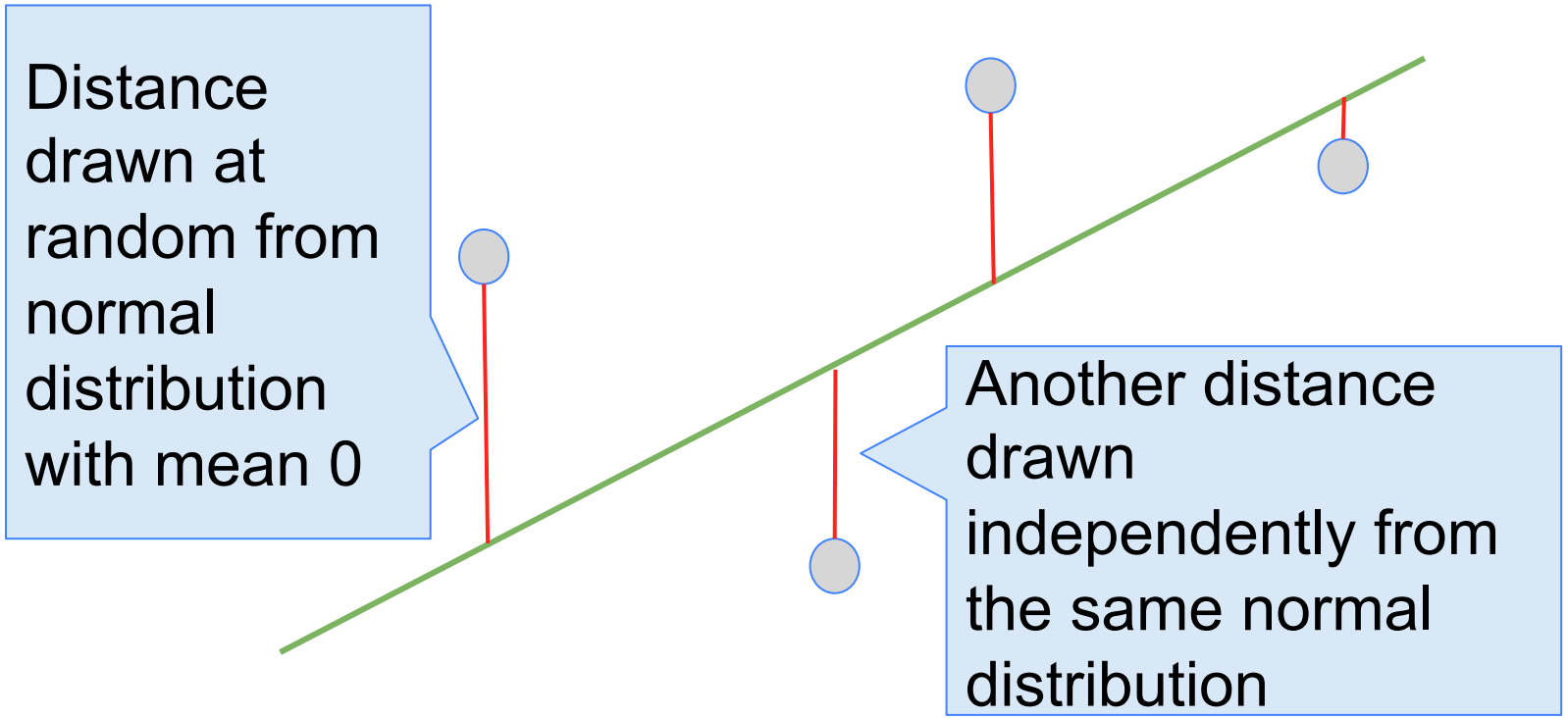
B. (SD of fitted) / (SD of y) = |r|

C. (SD of fitted) / (SD of residuals) = r

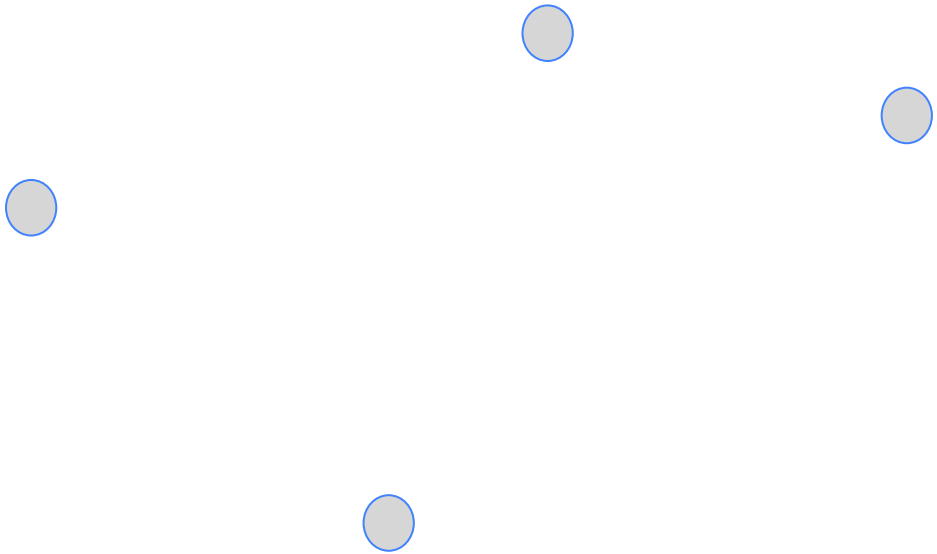
D. (SD of fitted) / (SD of residuals) = |r|

Regression Model

A “Model”: Signal + Noise



What We Get to See



(Demo)