Semana 13

Monday, October 25, 2021 2:11 PM

Residuals

A residual is the difference between a regression estimate and the actual value that we're trying to estimate.

- Errors in regression estimates
- There is one residual for each point (x,y)
- Residual = observed y - regresion estimate of y
  = observed y - height of regression line at x
  = vertical difference between point and line

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 linear regresion is appropriate.

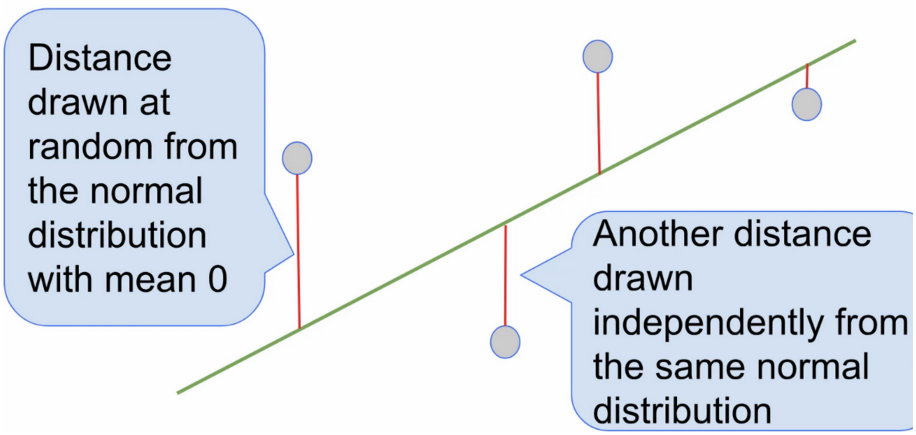
Residual variance

The mean of residual 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 residuals) + (variance of fitted values)

How does the SD of the fitted values relate to r?
(SD of fitted) / (SD of y) = |r|

Regression inference

A "model": Signal + noise



Regression prediction

If the data came from the regression model,

- The regression line is close to true line
- Given a new value of x, predict y by finding the point on the regression line at that x

Confidence interval for prediction

- Botstrap the scatter plot

- Get a prediction for y using the regression line that goes through the resampled plot
- Repeat the two steps above many times
- Draw the empirical histogram of all the predictions.
- Get the "middle 95%" interval.
- That’s an approximate 95% confidence interval for the predicted value of y

Confidence interval for true slope

- Bootsrap the scatter plot.
- Find the slope of the regression line through the bootstrapped plot.
- Repeat
- Draw the empirical histogram of all the generated slopes.
- Get the "middle 95%" interval.
- That’s an approximate 95% confidence interval for the slope of the true line.

Test whether there really is a slope

- Null hypothesis: The slope of the true line is 0
- Alternative hypothesis: No, its not.
- Method:
    - Construct a bootstrap confidence interval for the true slope.
    - If the interval doesn’t contain 0, reject the null hypothesis.
    - If the interval does contain 0, there isn't enough evidence to reject the null hypothesis.