Data 8, Lab 12

Residuals

Hubert Luo

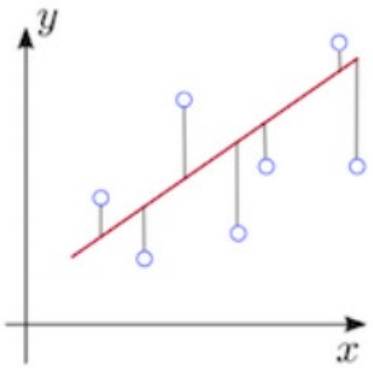
Fall 2019

5 December 2019

Berkeley
UNIVERSITY OF CALIFORNIA

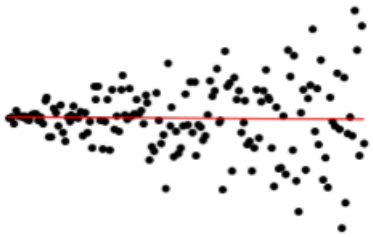
Regression

- Predict the value of a **continuous** random variable
  - Example: predict the height of a child given the height of a parent
- In Data 8, we use an algorithm called **Least Squares**
- This creates a linear prediction



Residuals

- Residual = Actual Value – Predicted Value (via Least Squares)
- The residual plot of a good regression (a linear model) shows no patterns in the graph of the residuals
- Average of residuals is always 0
- Heteroscedastic: Uneven variation of residuals around the horizontal line at 0
  - The regression estimates are not equally accurate
  - Example:



Standard Deviations

- SD of the Residuals = np.sqrt(1-r**2) * SD_of_y
- It is also a fraction of the SD of the response variable (y)
- The SD tells us how good the linear predictor
  - The smaller the SD of Residuals, the closer the residuals are to their mean
  - Mean of Residuals is always 0
  - Example: If r=1 (perfect correlation), there SD of residuals is 0 since we have a perfect linear relationship between X and y
- SD of Predicted Values = |r| * SD of y

Announcements

- Topical Review Labs for RRR Week
- Leads review on Wed and Fri of RRR Week
- Watch the past semesters Wagner Security lecture
  - Final question might be related
  - can be found on last semester's calendar: http://data8.org/sp19/ on Friday 4/19
- Course evaluations threshold is 80%. If the entire class meets the threshold, everyone gets an extra point on the final.
- If you're interested, apply to be a Lab Assistant next semester!
  - Application will be posted during winter break on piazza