

Homework 03

Brandon Hosley

University of Illinois - Springfield

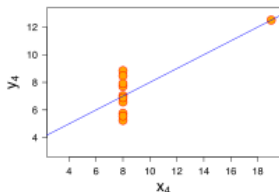
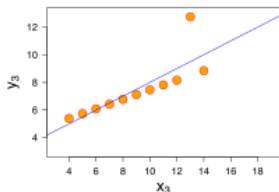
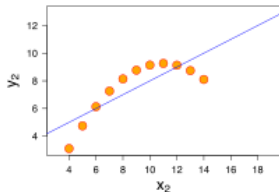
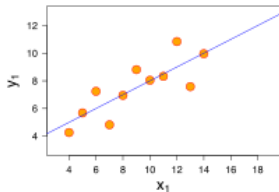
September 15, 2020

Overview

- 1 Q1: Linear Regression
- 2 Q2: Hastie and Tibshirani Summary

Linear Regression Problem 1: Simplicity

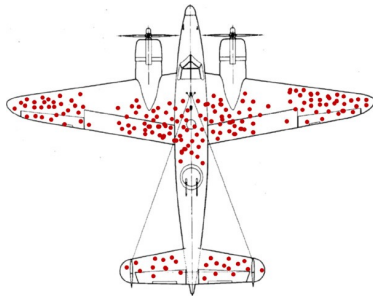
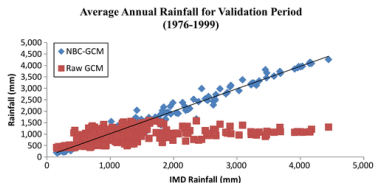
- + Computationally easy
 - Can only accurately represent simple relationships
 - Can only accurately represent linear relationships



Linear Regression Problem 2: Selection Bias

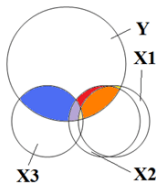
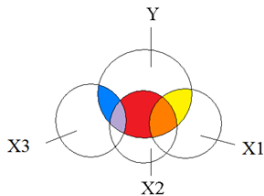
Linear regression is susceptible to selection bias

- A type of overfitting
- Occurs when a type of data is over-represented in test set



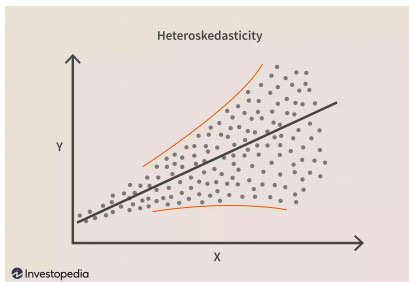
Linear Regression Problem 3: Multicollinearity

- When multiple predictors share a linear relationship
- Small changes are magnified in the model
- Heavily correlated predictors may cause redundancy in model



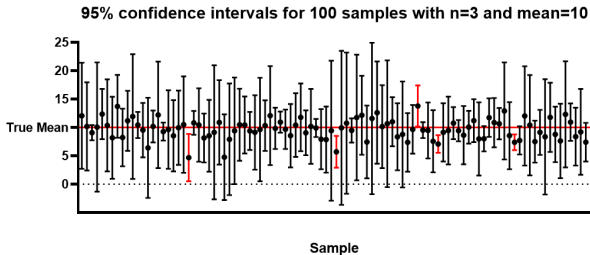
Linear Regression Problem 4: Heteroscedasticity

- Divergent data
- May be easily bound between two curves
- Difficult to accurately predict



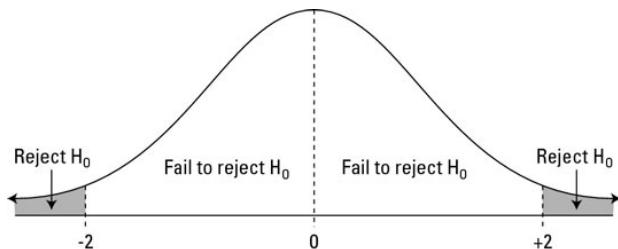
Tibshirani Lecture: Linear Regression

- Linear Regression
 - Simple approximation method
 - Great for estimating slope of data
 - From this one may generate confidence interval



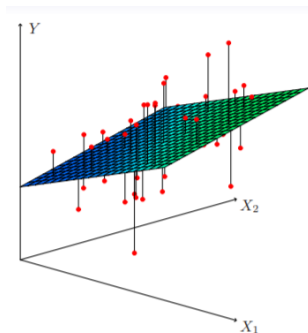
Tibshirani Lecture: Linear Regression

- Linear Regression
 - Simple approximation method
 - Great for estimating slope of data
 - From this one may generate confidence interval
- Hypothesis Testing
 - Testing against a null hypothesis
 H_0 : No relationship between X and Y .
 - Testing for the probability of independent variable distribution



Tibshirani Lectures: Multiple Linear Regression

- Multiple Linear Regression
 - Fitting data to a hyperplane instead of line
 - Works best when variables are independent



Tibshirani Lectures: Choosing Variables

- Selection from all subsets
- Forward Selection - adding variables with highest significance
- Backward Selection - removing variables with least significance