Simple Linear Regression and Least Squares

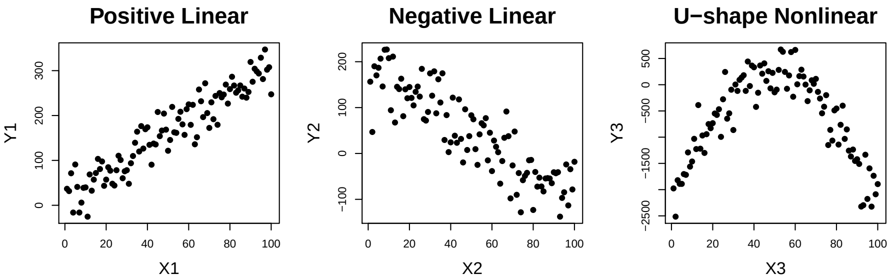
Christine P. Chai

Department of Statistical Science
Duke University

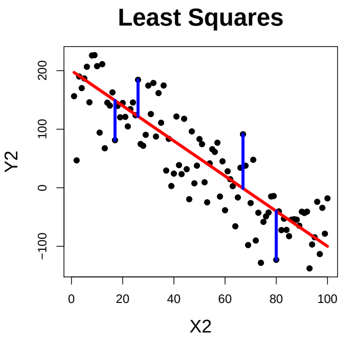
February 9, 2016

Simple Linear Regression

- Linear regression: $Y = \alpha + \beta X + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$
- Predict Y (response variable) from X (explanatory variable)
- e.g. predict a child's adult height using parents' heights
- The relationship between X and Y needs to follow a **linear** trend
- Error term ϵ is independent, has mean 0, and normally distributed



Least Squares Approach



- X is explanatory; Y is response
- Datapoints (x_i, y_i) for $i = 1, \cdots, n$
- Estimated $(x_i, \hat{y}_i)$ with $\hat{y}_i = \alpha + \beta x_i$
- Goal: Minimize $\sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$
  Sum of **squared distance** between estimated $\hat{Y}$ and real Y values
- Estimated slope

$$\hat{\beta} = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{x})^2} = \frac{Cov(X, Y)}{Var(X)}$$

- Estimated intercept $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$

Interpretation of Regression Line

- $Y = \alpha + \beta X + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$
- For every unit increase in X, Y **is expected to** change by β units
- α serves as an intercept – the value of Y when $X = 0$
- α may or may not have a meaning

- e.g. predict a person's salary (Y) using years of education (X)
- α is the expected salary for a person with 0 years of education
- It is possible for a person to have no education at all

- e.g. predict wives' age (Y) using husbands' age (X)
- $\alpha = Y$ when $X = 0$, but no one gets married at 0 years old
- Intercept: meaningless, simply adjusts the height of the line