

# Regression

## APAM E4990

### Modeling Social Data

Jake Hofman

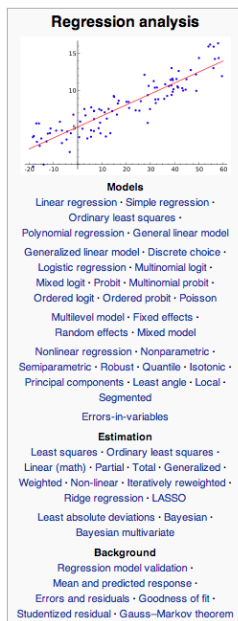
Columbia University

March 8, 2019

# Definition

?

# Definition



# Definition

*“The primary goal in a regression analysis is to **understand**, as far as possible **with the available data**, how the conditional distribution of the **response varies across subpopulations** determined by the possible values of the predictor or **predictors**.”*

- “Applied Regression Including Computing and Graphics”  
Cook & Weisberg (1999)

# Goals

## Describe

Provide a **compact summary** of outcomes under different conditions

## Predict

Make forecasts for **future** outcomes or **unobserved** conditions

## Explain

Account for **associations** between predictors and outcomes

# Goals

## Describe

Provide a **compact summary** of outcomes under different conditions

Never “false”, but may be wasteful or misleading

## Predict

Make forecasts for **future** outcomes or **unobserved** conditions

Varying degrees of success, often room for improvement

## Explain

Account for **associations** between predictors and outcomes

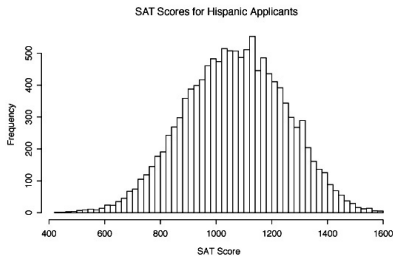
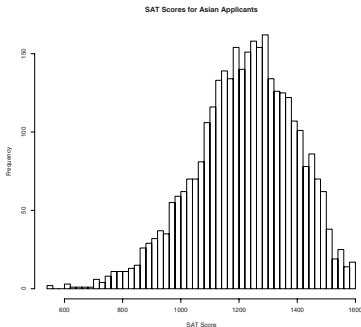
Difficult to establish causality in observational studies

See “Regression Analysis: A Constructive Critique”, Berk (2004)

# Goals

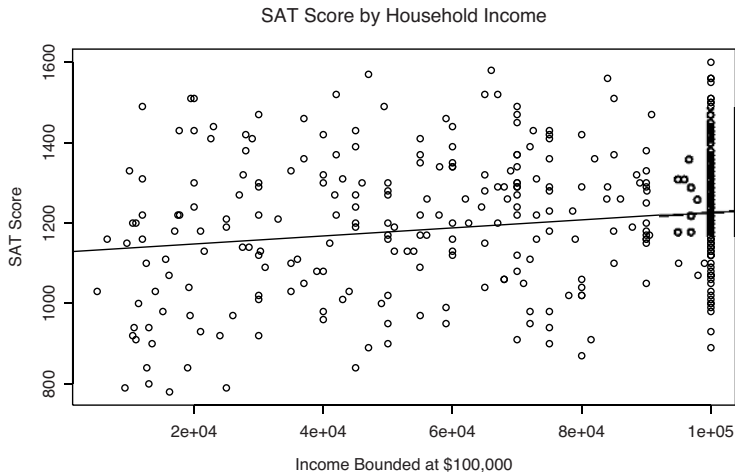
Models should be **flexible** enough to **describe observed** phenomena  
but **simple** enough to **generalize** to **future** observations

# Examples<sup>1</sup>



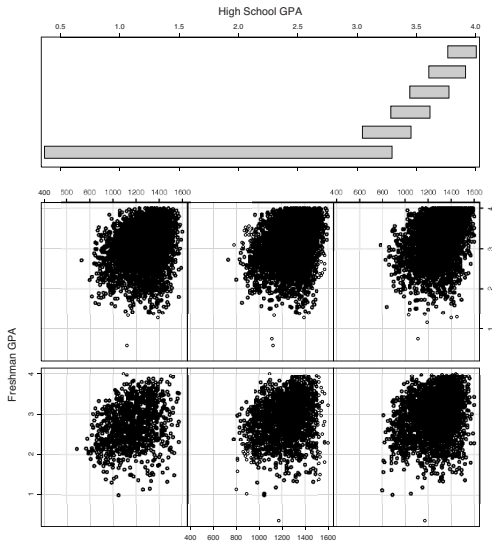


# Examples<sup>1</sup>



<sup>1</sup>“Statistical Learning from a Regression Perspective”, Berk (2008)

# Examples<sup>1</sup>



<sup>1</sup> "Statistical Learning from a Regression Perspective", Berk (2008)

# Framework

- Specify the **outcome** and **predictors**, along with the **form of the model** relating them
- Define a **loss function** that quantifies how close a model's predictions are to observed outcomes
- Develop an **algorithm** to fit the model to the observations by **minimizing this loss**
- **Assess** model performance and **interpret** results.