# Regression APAM E4990 Modeling Social Data

Jake Hofman

Columbia University

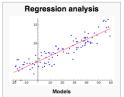
March 8, 2019

## **Definition**

?



#### Definition



Linear regression · Simple regression ·
Ordinary least squares ·
Polynomial regression · General linear model

Generalized linear model · Discrete choice ·
Logistic regression · Multinomial logit ·
Mixed logit · Probit · Multinomial probit ·
Ordered logit · Ordered probit · Poisson

Multilevel model • Fixed effects • Random effects • Mixed model Nonlinear regression • Nonparametric •

Semiparametric - Robust - Quantile - Isotonic Principal components - Least angle - Local Segmented

Errors-in-variables

#### Estimation

Least squares · Ordinary least squares ·
Linear (math) · Partial · Total · Generalized ·
Weighted · Non-linear · Iteratively reweighted ·
Ridge regression · LASSO

Least absolute deviations · Bayesian ·

#### Bayesian multivariate

#### Background

Regression model validation •
Mean and predicted response •
Errors and residuals • Goodness of fit •
Studentized residual • Gauss–Markov theorem

## 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

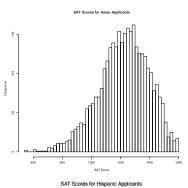
See "Regression Analysis: A Constructive Critique", Berk (2004)

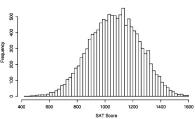
### Goals

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



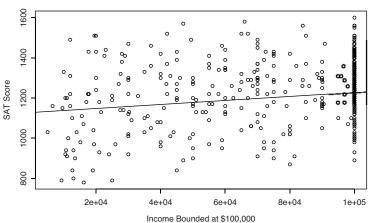
# $\mathsf{Examples}^1$





# Examples<sup>1</sup>

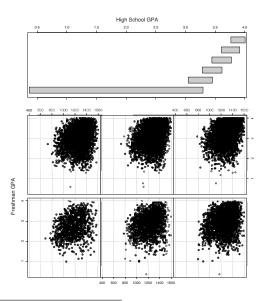




Jake Hofman (Columbia University)

<sup>1 &</sup>quot;Statistical Learning from a Regression Perspective", Berk (2008) 📳 👢 🔊 🤉

# Examples<sup>1</sup>



<sup>1 &</sup>quot;Statistical Learning from a Regression Perspective", Berk (2008) ≥ Section 2008

 ${\sf Jake\; Hofman} \quad {\sf (Columbia\; University)} \qquad \qquad {\sf Regression} \qquad \qquad {\sf March\; 8,\; 2019} \quad \ \, 5 \not - 6$ 

### 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.