Other considerations in regression analysis

- Outliers and influential observations.
- Model evaluation and comparison
- Model selection
- Hidden extrapolation
- Multiple testing *or* Multiple comparison

Extending linear regression

- **Linear regression can be extended to nonlinear regression using transformed variables such as X^2, log Y , etc.**

- **Generalized linear models (GLM) are linear regression models for non-Gaussian Y variables such as categorical variables.**

- **Local regression applies linear regression using observations close to individual X values.**

- **Regression models have also been extended to more complex types of Y variables.**

Concluding remarks

- Association patterns are everywhere.
- They represent information we can utilize
  - To explain
  - To estimate
  - To predict
- Association does not equal causation.
- Using a single set of data and search for various association patterns among a large number of variables is dangerous.
- Models, when used correctly, can be very useful.

Context Module

Descriptive analytics of text

- How to process text to generate numerical features

- How does modeling reveal patterns in the data