Coordinates

- Instructor: Merlise Clyde
  214A Old Chemistry
  Office Hours: TBA

- Teaching Assistants: Isaac Lavine

- Course: Theory and Application of linear models from both a frequentist (classical) and Bayesian perspective

- Prerequisites: linear algebra and a mathematical statistics course covering likelihoods and distribution theory (normal, t, F, chi-square, gamma distributions)

- Introduce R programming as needed

- Introduce Bayesian methods, but assume that you are co-registered in 601 or have taken it previously

more info on Course Website
http://stat.duke.edu/courses/Fall17/sta721
Build “regression” models that relate a response variable to a collection of covariates

- Goals of Analysis?
  - Predictive models
  - Causal interpretation
  - Testing of hypotheses
  - confirmatory or validation analyses

- Observational versus Experimental data? (Confounding)

- Sampling Schemes Generalizibility

- Statistical Theory
Prostate Example

- log PSA vs. log (can vol)
- log PSA vs. log (weight)
- log PSA vs. age
- log PSA vs. log (BPH)
- log PSA vs. SVI
- log PSA vs. log (cap pen)
- log PSA vs. gleason
- log PSA vs. PGS45
Simple Linear Regression

Simple Linear Regression:

\[ y_i = \beta_0 + \beta_1 x_i + \epsilon_i \text{ for } i = 1, \ldots, n \]

Rewrite in vectors:

\[
\begin{bmatrix}
  y_1 \\
  \vdots \\
  y_n
\end{bmatrix} = 
\begin{bmatrix}
  1 \\
  \vdots \\
  1
\end{bmatrix} \beta_0 + 
\begin{bmatrix}
  x_1 \\
  \vdots \\
  x_n
\end{bmatrix} \beta_1 + 
\begin{bmatrix}
  \epsilon_1 \\
  \vdots \\
  \epsilon_n
\end{bmatrix}
\]

\[
\begin{bmatrix}
  y_1 \\
  \vdots \\
  y_n
\end{bmatrix} = 
\begin{bmatrix}
  1 & x_1 \\
  \vdots & \vdots \\
  1 & x_n
\end{bmatrix} 
\begin{bmatrix}
  \beta_0 \\
  \beta_1
\end{bmatrix} + 
\begin{bmatrix}
  \epsilon_1 \\
  \vdots \\
  \epsilon_n
\end{bmatrix}
\]

\[\mathbf{Y} = \mathbf{X}\mathbf{\beta} + \mathbf{\epsilon}\]
Multiple Regression

\[ y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \epsilon_i \]

Design matrix

\[
X = \begin{bmatrix}
1 & x_{11} & \ldots & x_{p1} \\
1 & x_{12} & \ldots & x_{p2} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_{1n} & \ldots & x_{pn}
\end{bmatrix}
\]

\[ Y = X\beta + \epsilon \]

what should go into \( X \) and do we need all columns of \( X \) for inference about \( Y \)?
Nonlinear Models

Regression function may be an intrinsically nonlinear function of $t$

$$E[Y_i] = f(t_i, \theta)$$
Quadratic Linear Regression

Taylor’s Theorem:

\[ f(t_i, \theta) = f(t_0, \theta) + (t_i - t_0)f'(t_0, \theta) + (t_i - t_0)^2 \frac{f''(t_0, \theta)}{2} + R(t_i, \theta) \]

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i \text{ for } i = 1, \ldots, n \]

Rewrite in vectors:

\[
\begin{bmatrix}
  y_1 \\
  \vdots \\
  y_n
\end{bmatrix}
= \begin{bmatrix}
  1 & x_1 & x_1^2 \\
  \vdots & \vdots & \vdots \\
  1 & x_n & x_n^2
\end{bmatrix}
\begin{bmatrix}
  \beta_0 \\
  \beta_1 \\
  \beta_2
\end{bmatrix}
+ \begin{bmatrix}
  \epsilon_1 \\
  \vdots \\
  \epsilon_n
\end{bmatrix}
\]

\[ Y = X\beta + \epsilon \]

Quadratic in \( x \), but linear in \( \beta \)’s, but remainder term is in errors \( \epsilon \)
Polynomial Linear Regression

Polynomial Regression:

\[ y_i = \sum_{j=0}^{q} \beta_j x_i^j + \epsilon_i \text{ for } i = 1, \ldots, n \]

Rewrite in vector notation:

\[
\begin{bmatrix}
    y_1 \\
    \vdots \\
    y_n 
\end{bmatrix} =
\begin{bmatrix}
    1 & x_1 & x_1^2 & \ldots & x_1^q \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    1 & x_n & x_n^2 & \ldots & x_n^q 
\end{bmatrix}
\begin{bmatrix}
    \beta_0 \\
    \beta_1 \\
    \beta_2 \\
    \vdots \\
    \beta_q 
\end{bmatrix} +
\begin{bmatrix}
    \epsilon_1 \\
    \vdots \\
    \epsilon_n 
\end{bmatrix}
\]

\[ Y = X\beta + \epsilon \]

How large should \( q \) be?

Use Nonlinear Regression or other Nonparametric models.
Kernel Regression:

\[ y_i = \beta_0 + \sum_{j=1}^{J} \beta_j e^{-\lambda(x_i-k_j)^d} + \epsilon_i \text{ for } i = 1, \ldots, n \]

where \( k_j \) are kernel locations and \( \lambda \) is a smoothing parameter.

\[
\begin{bmatrix}
y_1 \\
\vdots \\
y_n
\end{bmatrix} = 
\begin{bmatrix}
1 & e^{-\lambda(x_1-k_1)^d} & \ldots & e^{-\lambda(x_1-k_J)^d} \\
\vdots & \vdots & \ddots & \vdots \\
1 & e^{-\lambda(x_n-k_1)^d} & \ldots & e^{-\lambda(x_n-k_J)^d}
\end{bmatrix} 
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_J
\end{bmatrix} + 
\begin{bmatrix}
\epsilon_1 \\
\vdots \\
\epsilon_n
\end{bmatrix}
\]

\[
\begin{align*}
\mathbf{Y} &= \mathbf{X}\beta + \epsilon \\
\text{Linear in } \beta \text{ given } \lambda \\
\text{Learn } \lambda \text{ and } J
\end{align*}
\]
Kernel Regression Example

The diagram illustrates the output of a kernel regression analysis on a dataset. The x-axis represents time in milliseconds (ms), and the y-axis represents acceleration in g-force. The graph shows the relationship between time and acceleration, with three key lines:

- **Fitted Mean**: Represented by a solid red line, this line indicates the mean acceleration at each time point.
- **95% Quantile**: Dashed black line, indicating the upper bound of the 95% confidence interval for the fitted mean.
- **5% Quantile**: Dashed grey line, indicating the lower bound of the 95% confidence interval for the fitted mean.

The data points are represented by grey circles, which scatter across the graph, showing the variability in acceleration at different time points.
Hierarchical Models - Spinal Bone Density

The diagram illustrates the relationship between age (in years) and spinal bone marrow density for different ethnic groups. The x-axis represents age, ranging from 10 to 25 years, and the y-axis represents spinal bone marrow density, ranging from 0.6 to 1.4. The four quadrants are color-coded: Hispanic, White, Asian, and Black. Each quadrant contains scatter plots showing the distribution of data points for different individuals within each ethnic group, indicating variations in bone marrow density across different ages.
Generic Linear Model

Generic Model in Matrix Notation is

\[ Y = X \beta + \epsilon \]

- \( Y \) \((n \times 1)\) vector of response (observe)
- \( X \) \((n \times p)\) design matrix (observe)
- \( \beta \) \((p \times 1)\) vector of coefficients (unknown)
- \( \epsilon \) \((n \times 1)\) vector of “errors” (unobservable)

Goals:
- What goes into \( X \)? (model building and model selection)
- What if several models are equally good? (model averaging or ensembles)
- What about the future? (Prediction)
- uncertainty quantification - assumptions about \( \epsilon \)

*All models are wrong, but some may be useful* (George Box)
Ordinary Least Squares

Goal: Find the best fitting “line” or “hyper-plane” that minimizes

\[ \sum (Y_i - x_i^T \beta)^2 = (Y - X\beta)^T (Y - X\beta) = \|Y - X\beta\|^2 \]

- Optimization problem
- May over-fit ⇒ add other criteria that provide a penalty “Penalized Least Squares”
- Robustness to extreme points ⇒ replace quadratic loss with other functions
- no notion of uncertainty of estimates
- no structure of problem (repeated measures on individual, randomization restrictions, etc)

Need Distribution Assumptions of \( Y \) (or \( \epsilon \)) for testing and uncertainty measures ⇒ Likelihood and Bayesian inference
Philosophy

- for many problems frequentist and Bayesian methods will give similar answers (more a matter of taste in interpretation)
- For small problems, Bayesian methods allow us to incorporate prior information which provides better calibrated answers
- For problems with complex designs and/or missing data Bayesian methods are often easier to implement (do not need to rely on asymptotics)
- For problems involving hypothesis testing or model selection frequentist and Bayesian methods can be strikingly different.
- Frequentist methods often faster (particularly with “big data”) so great for exploratory analysis and for building a “data-sense”
- Bayesian methods sit on top of Frequentist Likelihood

Important to understand advantages and problems of each perspective!