



Regression Models Overview

Jeffrey Leek
Johns Hopkins Bloomberg School of Public Health

Regression Models Content

- Linear regression
- Multiple Regression
- Confounding
- Residuals and diagnostics
- Prediction using linear models
- Model misspecification
- Scatterplot smoothing/splines
- Machine learning via regression
- Resampling inference in regression, bootstrapping, permutation tests
- Weighted regression
- Mixed models (random intercepts)

A historically famous idea, Regression to the Mean

- Why is it that the children of tall parents tend to be tall, but not as tall as their parents?
- Why do children of short parents tend to be short, but not as short as their parents?
- Why do parents of very short children, tend to be short, but not as short as their child? And the same with parents of very tall children?
- Why do the best performing athletes this year tend to do a little worse the following?

Basic regression model with additive Gaussian errors

- Least squares is an estimation tool, how do we do inference?
- Consider developing a probabilistic model for linear regression

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- Here the ϵ_i are assumed iid $N(0, \sigma^2)$.
- Note, $E[Y_i | X_i = x_i] = \mu_i = \beta_0 + \beta_1 x_i$
- Note, $\text{Var}(Y_i | X_i = x_i) = \sigma^2$.
- Likelihood equivalent model specification is that the Y_i are independent $N(\mu_i, \sigma^2)$.

Multivariable regression analyses

- An insurance company is interested in how last year's claims can predict a person's time in the hospital this year.
 - They want to use an enormous amount of data contained in claims to predict a single number. Simple linear regression is not equipped to handle more than one predictor.
- How can one generalize SLR to incorporate lots of regressors for the purpose of prediction?
- What are the consequences of adding lots of regressors?
 - Surely there must be consequences to throwing variables in that aren't related to Y ?
 - Surely there must be consequences to omitting variables that are?