Week 4

Logistic Regression

I. Odds
   a. A way to express likelihood event will take place
   b. Written as X/Y, X:Y, or X to Y
   c. Gambling odds called odds against
      i. Probability event will not happen
      ii. 10 to 1 odds == bet $1 and win get $10 + $1 (initial bet)
      iii. Probability an event will not happen > probability that it will happen
   d. Probability an event will happen > probability that it will not happen
   e. Also called odds for or odds on
   f. 2 to 1 odds means event twice as likely to happen than not
   g. Get back your initial bet + stake of $2??

II. Statistical Odds
   a. Viewed as ratio of probabilities
   b. Odds used in favor that event will happen over the probability that it will not happen
   c. If Odds for is 2:1 then:
      i. Odds(for) = 2/1 = p/(1-p) thus, 2(1-p) = p, 2-2p =p, 2=3p, 2/3 = p = 0.6667 or 66.67%
      ii. P is probability
      iii. P = Odds(for) / (1 + Odds(for))
      iv. Odds(for) =p/ (1-p)
      v. P = 12.5% or 0.125 so odds for is .125 / (1-.125) = 1/7

Logistic Regression

I. Linear regression can produce models with predicted values below 0 and above 1 making it unsuitable for binary responses
II. Use logistic regression to predict binary responses
III. Y is categorical (Yes/No), (Approve/Reject), (Pass/Fail) etc
IV. Result expressed as probability of being in a group
   a. Implying predicted value always between 0 and 1
V.