Logistic Regression

- Predicts the probability of poor care
  - Denote dependent variable "PoorCare" by y
  - $P(y = 1)$
- Then $P(y = 0) = 1 - P(y = 1)$
- Independent variables $x_1, x_2, \ldots, x_k$
- Uses the Logistic Response Function

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}}$$

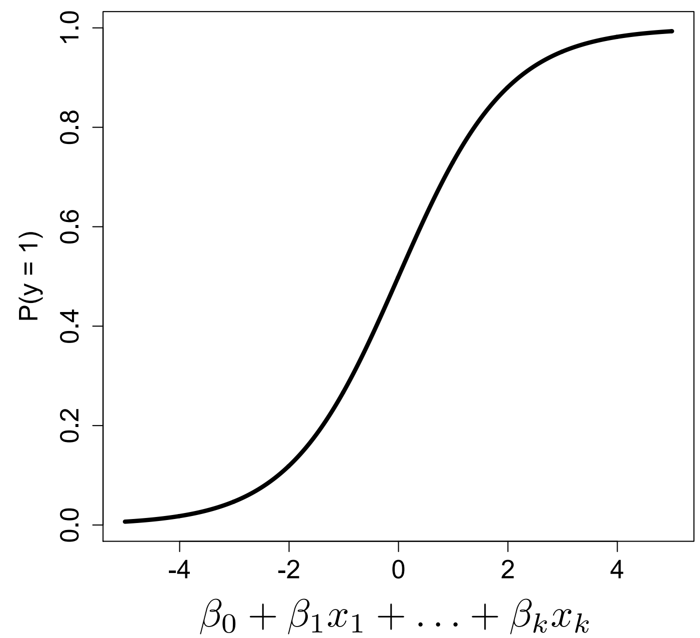
  - Nonlinear transformation of linear regression equation to produce number between 0 and 1

Poor Care = 1
Good Care = 0

Understanding the Logistic Function

$$P(y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k)}}$$

- Positive values are predictive of class 1
- Negative values are predictive of class 0



Understanding the Logistic Function

$$P(y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k)}}$$

- The coefficients are selected to
  - Predict a high probability for the poor care cases
  - Predict a low probability for the good care cases

Understanding the Logistic Function

$$P(y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}}$$

- We can instead talk about Odds (like in gambling)

$$\text{Odds} = \frac{P(y=1)}{P(y=0)}$$

- Odds > 1 if y = 1 is more likely
- Odds < 1 if y = 0 is more likely

The Logit

- It turns out that

$$\text{Odds} = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k}$$

$$log(\text{Odds}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

- This is called the "Logit" and looks like linear regression
- The bigger the Logit is, the bigger $P(y = 1)$