Logistic Regression

Sravani Vadlamani

QMB 3200: Advanced Quantitative Methods

11/17/2020

Logistic Regression

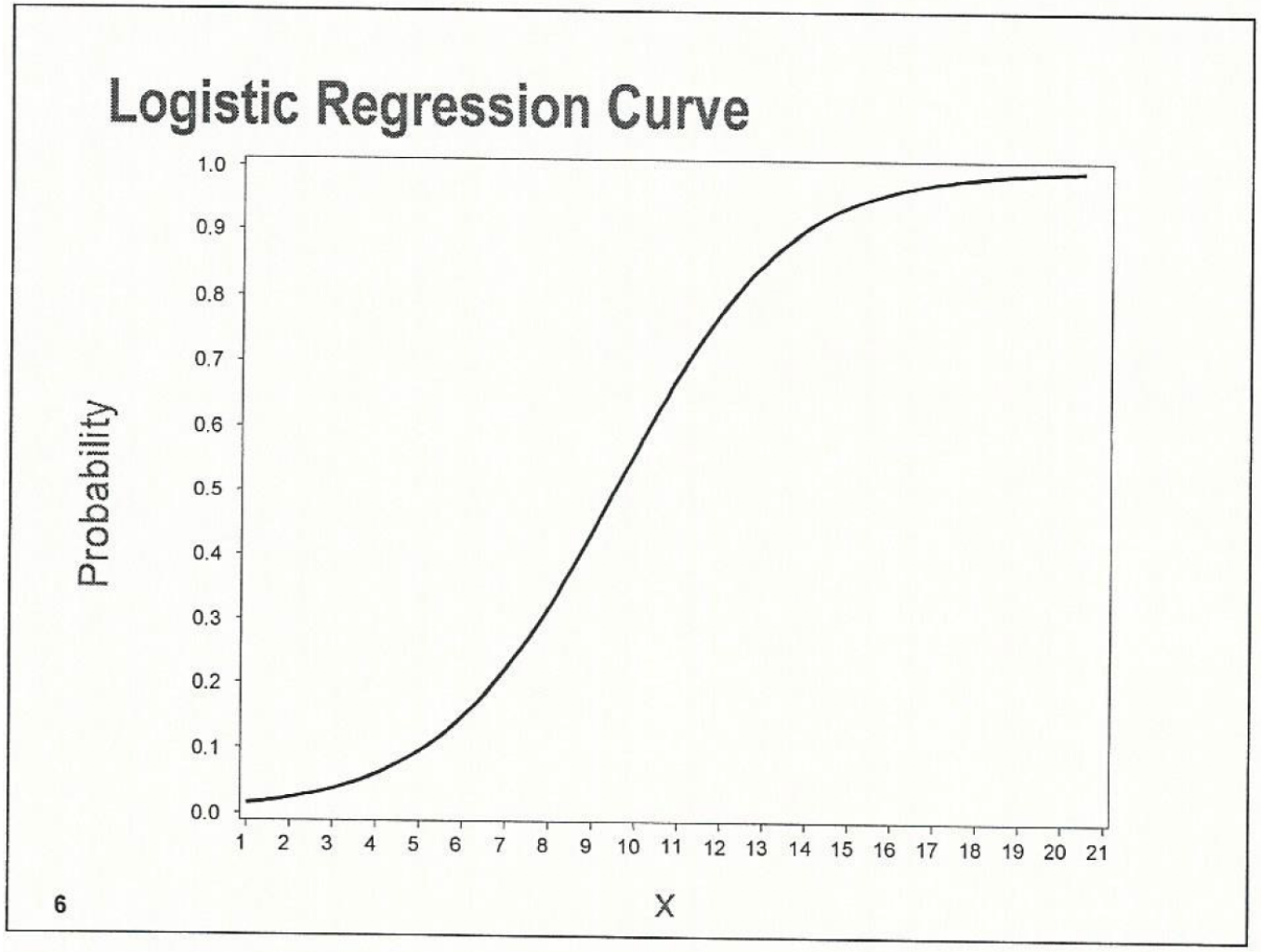
- An approximate regression model for binary data
- Examples
  - College Graduate (0 = no, 1 = yes)
  - Pass or Fail this class (0 = fail, 1 = pass)
  - Like football or not (1= like, 0 = not like)
- You model the probability of responding in each category

Logistic Regression

$$p = \beta_0 + \beta_1 X_1$$

- Issues
- Probabilities can take values between 0 and 1 (bounded constraint)
- Non-linear relation between probability and X variables

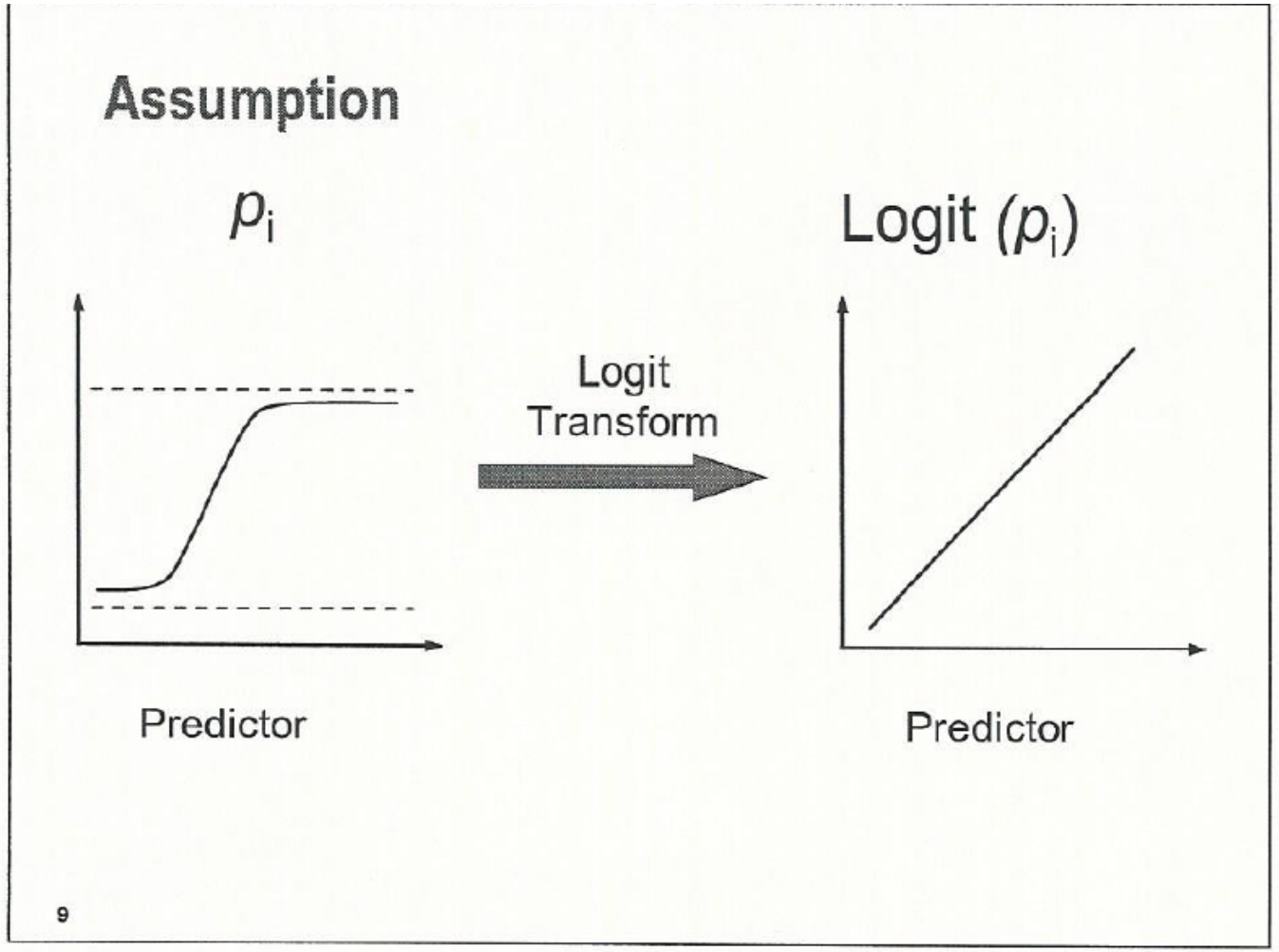
Logistic Regression



Logistic Regression

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})}}$$

Logistic Regression



Logit Regression to Logistic Regression

$$logit(P) = b_0 + b_i x_i$$

$$\ln\left(\frac{P}{1-P}\right) = b_0 + b_i x_i$$

$$\frac{P}{1-P} = e^{b_0 + b_i x_i}$$

$$P = \frac{e^{b_0 + b_i x_i}}{1+e^{b_0 + b_i x_i}}$$

Important Definitions

- P = probability of event
- Odds is the probability of event divided by the probability of non-event

$$Odds = \frac{P}{1 - P}$$

- Logit – Natural log of odds

$$\ln(Odds) = \ln\left(\frac{P}{1 - P}\right)$$

- In logistic regression, the log of the odds (logit) is linearly related to the predictor

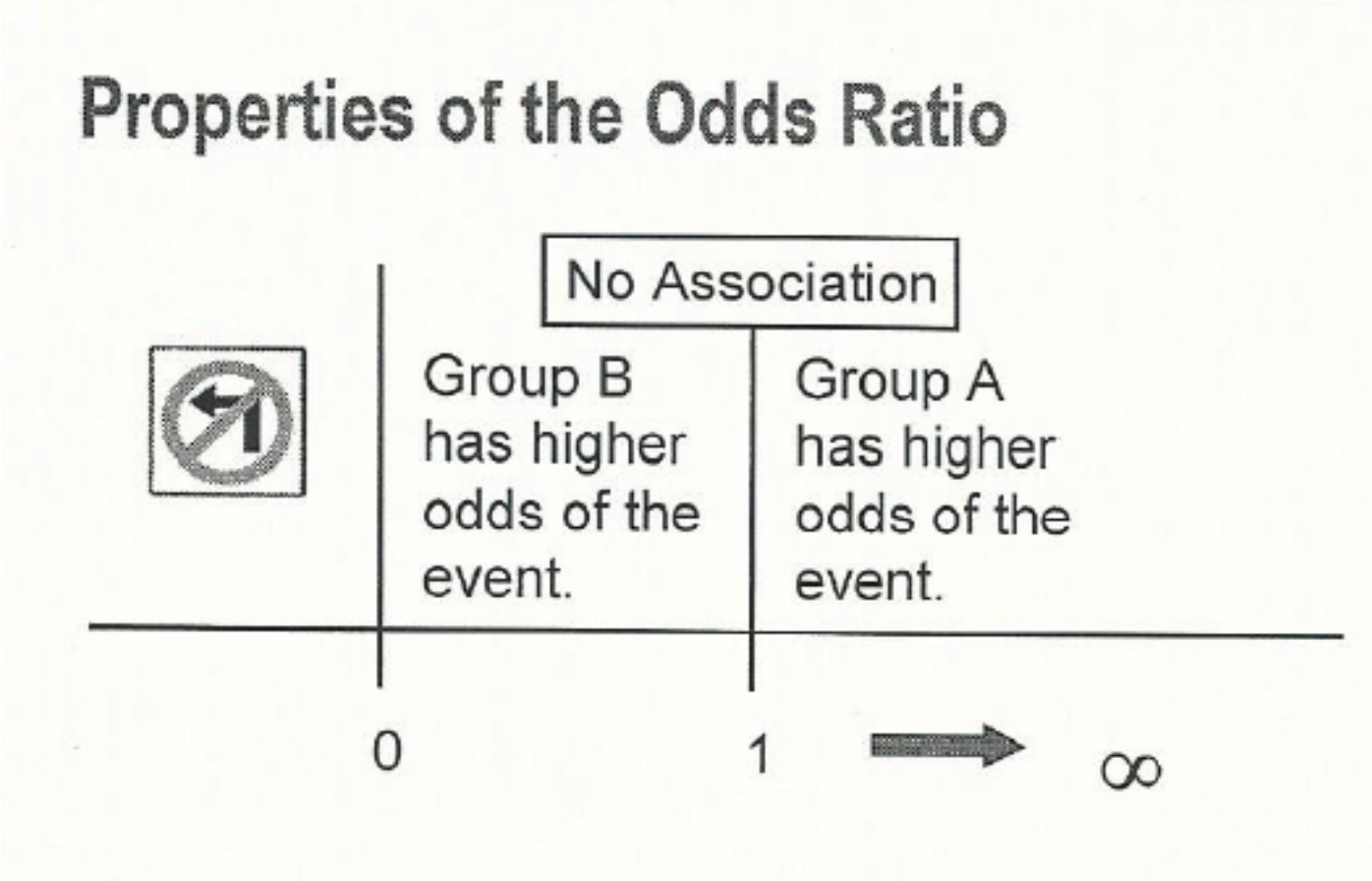
$$logit(P) = b_0 + b_i x_i$$

Odds Ratio

$$\text{Odds}_A = \frac{p_A}{1-p_A} \qquad \text{Odds}_B = \frac{p_B}{1-p_B}$$

$$\text{Odds Ratio} = \frac{\text{odds}_A}{\text{odds}_B}$$

Odds Ratio


Properties of the Odds Ratio
No Association
Group B has higher odds of the event.
Group A has higher odds of the event.
0
1
∞

Odds Ratio

Estimated logistic regression model:

$$\text{logit}(p) = -.7567 + .4373*(\texttt{gender})$$

Estimated odds ratio (Females to Males):

$$\text{odds ratio} = (e^{-.7567+.4373})/(e^{-.7567}) = 1.55$$