# Logistic Regression

Prof. Sharon McNicholas

STATS 780/CSE 780





# Introduction

- This lecture will cover logistic regression.
- Initially, we will look at binary logistic regression for one univariate predictor $X$.
- Then we will consider multiple predictors.
- Then, multinomial logistic regression.
- Note that Assignment 2 is now available, and concerns logistic regression.

# $\mathbb{E}[Y \mid x]$

- In simple linear regression,

$$\mathbb{E}[Y \mid x] = \beta_0 + \beta_1 x. \tag{1}$$

- From (1), we see that if $X \in \mathbb{R}$ then $Y$ can take any value on the real line, i.e, $Y \in \mathbb{R}$.

- But what happens if $Y$ is a binary (dichotomous) random variable, i.e., if $Y$ can only take values in $\{0, 1\}$?

- Then, $\mathbb{E}[Y \mid x]$ should be defined appropriately.





# Using the Logit Link

- For ease of notation, let $\pi(x) = \mathbb{E}[Y \mid x]$.
- We use the logit so that

$$\log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x,$$

i.e.,

$$\pi(x) = \frac{\exp\{\beta_0 + \beta_1 x\}}{1 + \exp\{\beta_0 + \beta_1 x\}}. \tag{2}$$

- From (3), it is clear that $\pi(x) \in (0, 1)$.
- Important note: if $Y$ is binary, i.e., taking values in $\{0, 1\}$, then

$$\mathbb{E}[Y] = 0\mathbb{P}[Y = 0] + 1\mathbb{P}[Y = 1] = \mathbb{P}[Y = 1],$$

so $\pi(x)$ is the same as $\mathbb{P}[Y = 1 \mid x]$.

# The Model

- Sometimes, people describe the binary logistic regression model as

$$\log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x.$$

- However, this is not a statistical model because there is no error term.
- Recall that the simple linear regression model is

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

- Note that another way of writing this is

$$Y = \mathbb{E}[Y \mid X] + \epsilon.$$





# The Model contd.

- Similarly, we can write the logistic regression model as

$$Y = \mathbb{E}[Y \mid X] + \epsilon,$$

i.e., as

$$Y = \pi(X) + \epsilon.$$

- But what is the distribution of $\epsilon$ here?
- If $Y = 1$, then $\epsilon = 1 - \pi(X)$ with probability $\pi(X)$.
- If $Y = 0$, then $\epsilon = -\pi(X)$ with probability $1 - \pi(X)$.
- That is, $\epsilon$ follows a distribution with mean $0$ and variance $\pi(X)[1 - \pi(X)]$.

# The Model contd.

- So $Y \mid x$ follows a distribution with mean $\pi(x)$ and variance $\pi(x)[1-\pi(x)]$.

- That is, $Y \mid x$ follows a Bernoulli distribution with success probability $\pi(x)$.

- And we know the relationship with the binomial distribution...





# The Odds Ratio

- Suppose that $X$ is also binary.

- The odds of the outcome, i.e., $Y=1$, when $x=1$ is $\pi(1)/[1-\pi(1)]$.

- The odds of the outcome, i.e., $Y=1$, when $x=0$ is $\pi(0)/[1-\pi(0)]$.

- The odds ratio is just the ratio of these odds:
$$\text{Odds Ratio} = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}.$$

- We can show that
$$\text{Odds Ratio} = \exp\{\beta_1\} = e^{\beta_1}.$$

# Comments

- I know that there is a lot to take in here.
- As usual, background reading is important (more on this later).
- Now is a good time to look at an example.
- Then, we will move to multiple predictors.





# Multiple (Binary) Logistic Regression

- Now, we move to multiple binary logistic regression.
- We did this with linear models and GLMs, i.e., we considered multiple predictor variables (i.e., multiple independent variables) $X_1, \ldots, X_p$.
- We will allow flexibility in the type (continuous, binary, other categorical) of these predictor variables, e.g., $X_1$ need not be of the same type as $X_2$.
- For now, we are still be looking at binary $Y$.

# Multiple (Binary) Logistic Reg. contd.

- The model can be considered as arising from

$$\log\left[\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p,$$

  i.e.,

$$\pi(\mathbf{x}) = \frac{\exp\{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p\}}. \tag{3}$$

- As before,

$$\pi(\mathbf{x}) = \mathbb{E}[Y \mid \mathbf{x}] = \mathbb{P}[Y = 1 \mid \mathbf{x}].$$

- How do we decide whether to remove a predictor variable?



STATS 780/CSE 780 Prof. Sharon McNicholas

# Deviance

- In general, the deviance is given by

$$D = -2\log\left[\frac{\mathcal{L}_{\text{fitted model}}}{\mathcal{L}_{\text{saturated model}}}\right].$$

- This is minus twice the log of the likelihood ratio.

- When $Y$ is binary, $\mathcal{L}_{\text{saturated model}} = 1$ and so

$$D = -2\log\mathcal{L}_{\text{fitted model}}.$$

# Assessing Variable Importance

- The change in $D$ due to the inclusion of a predictor variable is

$$
\begin{aligned}
G &= D(\text{model without the variable}) - D(\text{model with the variable})\\
&= -2\log\left[\frac{\mathcal{L}_{\text{without the variable}}}{\mathcal{L}_{\text{with the variable}}}\right].
\end{aligned}
$$

- This is minus twice the log of the likelihood ratio.

- In general, $G \sim \chi^2_\nu$, where the degrees of freedom $\nu$ is the difference in the number of free parameters in the models (note that it will matter whether the variables are continuous or categorical and whether $Y$ is binary or multinomial).





# Multinomial Logistic Regression

- Now, we move to multinomial logistic regression (MLR).

- We proceed in a similar fashion to the binary logistic regression case.

- This time, however, $Y$ can have more than two categories.

- For the purposes of explanation, let us assume that the categories of $Y$ are coded 0, 1, or 2.

- Once the extension to three categories is understood, extension to more categories is straightforward.

# MLR: The Model

- Similarly to before, the model can be considered as arising from

$$\log\left[\frac{\mathbb{P}[Y=1\mid\mathbf{x}]}{\mathbb{P}[Y=0\mid\mathbf{x}]}\right]=\beta_{10}+\beta_{11}x_1+\cdots+\beta_{1p}x_p,$$

and

$$\log\left[\frac{\mathbb{P}[Y=2\mid\mathbf{x}]}{\mathbb{P}[Y=0\mid\mathbf{x}]}\right]=\beta_{20}+\beta_{21}x_1+\cdots+\beta_{2p}x_p.$$

- Here, $Y=0$ is the reference (baseline) category (outcome).





# MLR: The Model contd.

- It follows that

$$\begin{aligned}\pi_0(\mathbf{x})&=\mathbb{P}[Y=0\mid\mathbf{x}]\\&=\frac{1}{1+\exp\{\beta_{10}+\beta_{11}x_1+\cdots+\beta_{1p}x_p\}+\exp\{\beta_{20}+\beta_{21}x_1+\cdots+\beta_{2p}x_p\}}\end{aligned}$$

$$\begin{aligned}\pi_1(\mathbf{x})&=\mathbb{P}[Y=1\mid\mathbf{x}]\\&=\frac{\exp\{\beta_{10}+\beta_{11}x_1+\cdots+\beta_{1p}x_p\}}{1+\exp\{\beta_{10}+\beta_{11}x_1+\cdots+\beta_{1p}x_p\}+\exp\{\beta_{20}+\beta_{21}x_1+\cdots+\beta_{2p}x_p\}}\end{aligned}$$

$$\begin{aligned}\pi_2(\mathbf{x})&=\mathbb{P}[Y=2\mid\mathbf{x}]\\&=\frac{\exp\{\beta_{20}+\beta_{21}x_1+\cdots+\beta_{2p}x_p\}}{1+\exp\{\beta_{10}+\beta_{11}x_1+\cdots+\beta_{1p}x_p\}+\exp\{\beta_{20}+\beta_{21}x_1+\cdots+\beta_{2p}x_p\}}\end{aligned}$$

# MLR: Generalizing

- Generalizing to more than three categories for $Y$ is straightforward.
- For $C$ categories, coded as $0, 1, \ldots, C$, we have

$$\log\left[\frac{\mathbb{P}[Y = c \mid \mathbf{x}]}{\mathbb{P}[Y = 0 \mid \mathbf{x}]}\right] = \beta_{c0} + \beta_{c1}x_1 + \cdots + \beta_{cp}x_p$$

and

$$\pi_c(\mathbf{x}) = \mathbb{P}[Y = c \mid \mathbf{x}] = \frac{\exp\{\beta_{c0} + \beta_{c1}x_1 + \cdots + \beta_{cp}x_p\}}{1 + \sum_{k=1}^{C}\exp\{\beta_{k0} + \beta_{k1}x_1 + \cdots + \beta_{kp}x_p\}},$$

for $c = 1, \ldots, C$, where $\pi_0(\mathbf{x})$ is as defined before.





# MLR: Odds Ratio

- Hosmer and Lemeshow (2000)[a] give a formula like the following for the odds ratio of $Y = j$ versus $Y = 0$ for predictor variable value $x = a$ versus $x = b$.

$$\text{Odds Ratio}_j(a, b) = \frac{\mathbb{P}[Y = j \mid x = a]/\mathbb{P}[Y = 0 \mid x = a]}{\mathbb{P}[Y = j \mid x = b]/\mathbb{P}[Y = 0 \mid x = b]}.$$

- Formulae like this are useful but I think the best way to grasp this is through some examples...

[a]Hosmer, D.W. and Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wiley.

# Comments

- Now, we move on to examples.

- I strongly encourage you to try out some examples for yourself.

- For background reading, Faraway (2016) does some logistic regression but my favourite book is Hosmer and Lemeshow (2000).

- Next week, we will move to classification... in some respects, the move has already begun!