Logistic Regression

STA 2101 Fall 2019

See last slide for copyright information

Binary outcomes are common and important

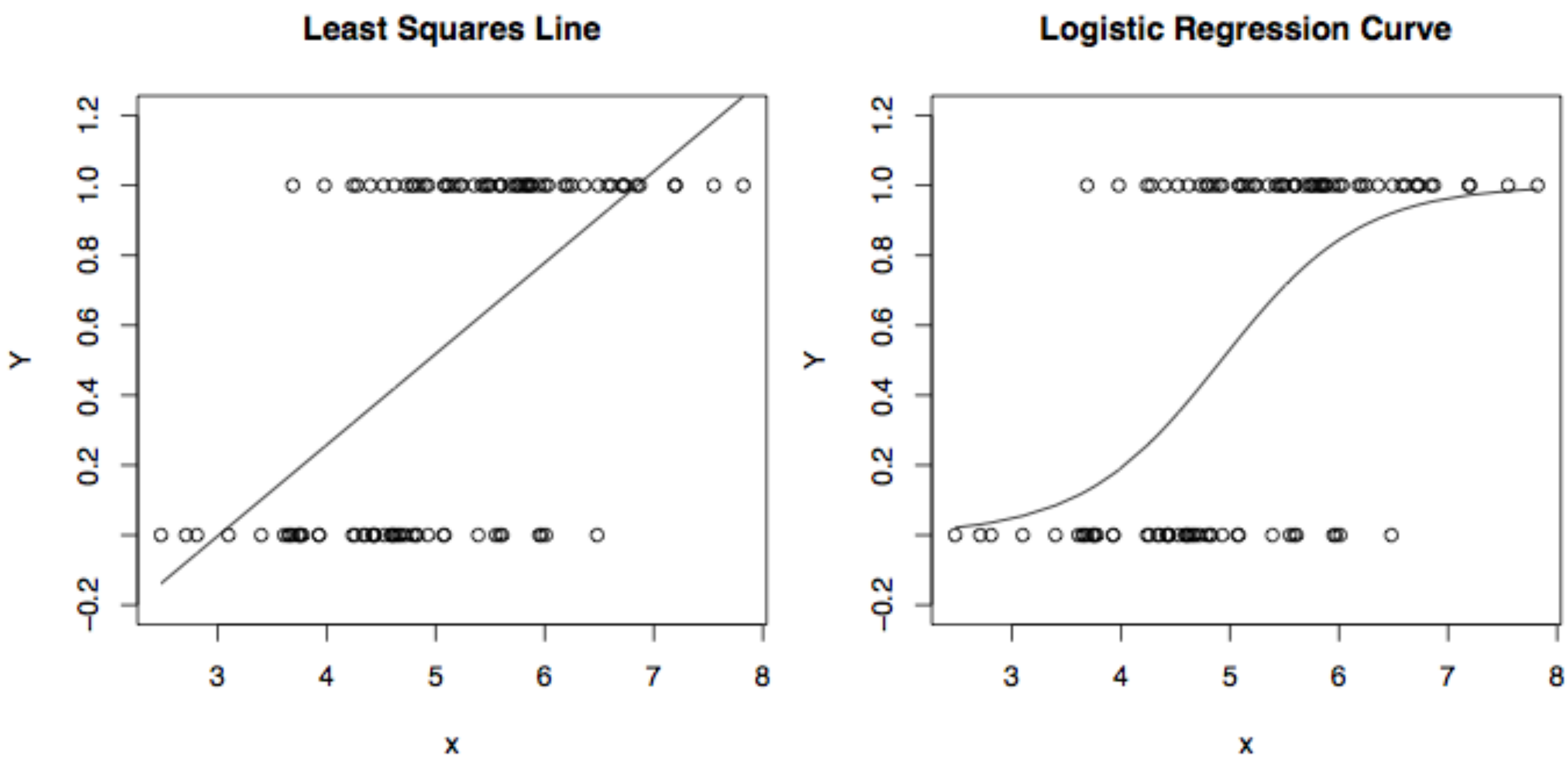
- The patient survives the operation, or does not.
- The accused is convicted, or is not.
- The customer makes a purchase, or does not.
- The marriage lasts at least five years, or does not.
- The student graduates, or does not.

Logistic Regression

Response variable is binary (Bernoulli):
1=Yes, 0=No

$$Pr\{Y = 1|\mathbf{X} = \mathbf{x}\} = E(Y|\mathbf{X} = \mathbf{x}) = \pi$$

Least Squares vs. Logistic Regression



The logistic regression curve arises from an indirect representation of the probability of Y=1 for a given set of x values.

Representing the probability of an event by π

$$\text{Odds} = \frac{\pi}{1-\pi}$$

$$\text{Odds} = \frac{\pi}{1-\pi}$$

- If P(Y=1)=1/2, odds = .5/(1-.5) = 1 (to 1)
- If P(Y=1)=2/3, odds = 2 (to 1)
- If P(Y=1)=3/5, odds = (3/5)/(2/5) = 1.5 (to 1)
- If P(Y=1)=1/5, odds = .25 (to 1)

The higher the probability, the greater the odds

$$\text{Odds} = \frac{\pi}{1 - \pi}$$

$$0 \leq \text{ Odds } < \infty$$

Linear regression model for the log odds of the event Y=1

for $i = 1, \ldots, n$

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}$$

Note π is a *conditional* probability.

Equivalent Statements

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}$$

$$\begin{aligned}\frac{\pi}{1-\pi} &= e^{\beta_0+\beta_1 x_1+\ldots+\beta_{p-1} x_{p-1}} \\ &= e^{\beta_0} e^{\beta_1 x_1} \cdots e^{\beta_{p-1} x_{p-1}},\end{aligned}$$

$$\pi = \frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_{p-1} x_{p-1}}}{1+e^{\beta_0+\beta_1 x_1+\ldots+\beta_{p-1} x_{p-1}}}.$$

$$E(Y|\mathbf{x}) = \pi = \frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_{p-1}x_{p-1}}}{1+e^{\beta_0+\beta_1 x_1+\ldots+\beta_{p-1}x_{p-1}}}$$

- A distinctly non-linear function
- Non-linear in the betas
- So logistic regression is an example of *non-linear regression*.

$F(x) = \frac{e^x}{1+e^x}$ is called the *logistic distribution.*

- Could use any cumulative distribution function:

$$\pi = F(\beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1})$$

- CDF of the standard normal used to be popular
- Called probit analysis
- Can be closely approximated with a logistic regression.

In terms of log odds, logistic regression is like regular regression

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}$$

In terms of plain odds,

- (Exponential function of) the logistic regression coefficients are *odds ratios*
- For example, "Among 50 year old men, the odds of being dead before age 60 are three times as great for smokers."

$$\frac{\text{Odds of death given smoker}}{\text{Odds of death given nonsmoker}} = 3$$

Logistic regression

- X=1 means smoker, X=0 means non-smoker
- Y=1 means dead, Y=0 means alive

- Log odds of death = $\beta_0 + \beta_1 x$

- Odds of death = $e^{\beta_0} e^{\beta_1 x}$

$$\text{Odds of Death} = e^{\beta_0} e^{\beta_1 x}$$

Group	x	**Odds of Death**
Smokers	1	$e^{\beta_0} e^{\beta_1}$
Non-smokers	0	e^{β_0}

$$\frac{\text{Odds of death given smoker}}{\text{Odds of death given nonsmoker}} = \frac{e^{\beta_0} e^{\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Cancer Therapy Example

$$\text{Log Survival Odds} = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_3 x$$

Treatment	d_1	d_2	**Odds of Survival** $= e^{\beta_0} e^{\beta_1 d_1} e^{\beta_2 d_2} e^{\beta_3 x}$
Chemotherapy	1	0	$e^{\beta_0} e^{\beta_1} e^{\beta_3 x}$
Radiation	0	1	$e^{\beta_0} e^{\beta_2} e^{\beta_3 x}$
Both	0	0	$e^{\beta_0} e^{\beta_3 x}$

x is severity of disease

For any given disease severity x,

$$\frac{\text{Survival odds with Chemo}}{\text{Survival odds with Both}} = \frac{e^{\beta_0} e^{\beta_1} e^{\beta_3 x}}{e^{\beta_0} e^{\beta_3 x}} = e^{\beta_1}$$

In general,

- When x_k is increased by one unit and all other explanatory variables are held constant, the odds of Y=1 are multiplied by e^{β_k}
- That is, e^{β_k} is an **odds ratio** --- the ratio of the odds of Y=1 when x_k is increased by one unit, to the odds of Y=1 when everything is left alone.
- As in ordinary regression, we speak of "controlling" for the other variables.

The conditional probability of Y=1

$$\pi_i = \frac{e^{\beta_0+\beta_1 x_{i,1}+\ldots+\beta_{p-1} x_{i,p-1}}}{1+e^{\beta_0+\beta_1 x_{i,1}+\ldots+\beta_{p-1} x_{i,p-1}}}$$

$$= \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1+e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}$$

This formula can be used to calculate a predicted P(Y=1|**x**). Just replace betas by their estimates

It can also be used to calculate the probability of getting the sample data values we actually did observe, as a function of the betas.

Likelihood Function

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} P(Y_i = y_i|\mathbf{x}_i) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i} \\
&= \prod_{i=1}^{n} \left(\frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1+e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}\right)^{y_i} \left(1 - \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1+e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}\right)^{1-y_i} \\
&= \prod_{i=1}^{n} \left(\frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1+e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}\right)^{y_i} \left(\frac{1}{1+e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}\right)^{1-y_i} \\
&= \prod_{i=1}^{n} \frac{e^{y_i \mathbf{x}_i^\top \boldsymbol{\beta}}}{1+e^{\mathbf{x}_i^\top \boldsymbol{\beta}}} \\
&= \frac{e^{\sum_{i=1}^{n} y_i \mathbf{x}_i^\top \boldsymbol{\beta}}}{\prod_{i=1}^{n}\left(1+e^{\mathbf{x}_i^\top \boldsymbol{\beta}}\right)}
\end{aligned}$$

Maximum likelihood estimation

- Likelihood = Conditional probability of getting the data values we did observe,
- As a function of the betas
- Maximize the (log) likelihood with respect to betas.
- Maximize numerically ("Iteratively re-weighted least squares")
- Likelihood ratio, Wald tests as usual
- Divide regression coefficients by estimated standard errors to get Z-tests of H_0: β_j=0.
- These Z-tests are like the t-tests in ordinary regression.

Copyright Information

This slide show was prepared by Jerry Brunner, Department of Statistics, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. These Powerpoint slides are available from the course website:
http://www.utstat.toronto.edu/brunner/oldclass/2101f19