Logistic Regression
Part One

STA2101/442 F 2012

See last slide for copyright information
Logistic Regression

For a binary dependent variable: 1=Yes, 0=No

\[
Pr\{Y = 1|X = x\} = \pi
\]
Least Squares vs. Logistic Regression

Least Squares Line

Logistic Regression Curve
Linear regression model for the log odds of the event $Y=1$

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

Note $\pi$ is a conditional probability.
Equivalent Statements

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

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

\[ = e^{\beta_0} e^{\beta_1 x_1} \ldots e^{\beta_{p-1} x_{p-1}} , \]

\[ \pi = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}}} . \]
A distinctly non-linear function

Non-linear in the betas

So logistic regression is an example of non-linear regression.
\[ F(x) = \frac{e^x}{1+e^x} \text{ is called the logistic distribution.} \]

- 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 = $e^{\beta_0} e^{\beta_1 x}$
Odds of Death = \( e^{\beta_0} e^{\beta_1 x} \)

<table>
<thead>
<tr>
<th>Group</th>
<th>( x )</th>
<th>Odds of Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smokers</td>
<td>1</td>
<td>( e^{\beta_0} e^{\beta_1} )</td>
</tr>
<tr>
<td>Non-smokers</td>
<td>0</td>
<td>( e^{\beta_0} )</td>
</tr>
</tbody>
</table>

\[
\frac{\text{Odds of death given smoker}}{\text{Odds of death given nonsmoker}} = \frac{e^{\beta_0} e^{\beta_1}}{e^{\beta_0}} = e^{\beta_1}
\]
Cancer Therapy Example

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

<table>
<thead>
<tr>
<th>Treatment</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>Odds of Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemotherapy</td>
<td>1</td>
<td>0</td>
<td>$e^{\beta_0} e^{\beta_1} e^{\beta_2} e^{\beta_3 x}$</td>
</tr>
<tr>
<td>Radiation</td>
<td>0</td>
<td>1</td>
<td>$e^{\beta_0} e^{\beta_2} e^{\beta_3 x}$</td>
</tr>
<tr>
<td>Both</td>
<td>0</td>
<td>0</td>
<td>$e^{\beta_0} e^{\beta_3 x}$</td>
</tr>
</tbody>
</table>

$x$ 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 independent variables are held constant, the odds of $Y=1$ are multiplied by $e^{\beta_k}$.
- That is, $e^{\beta_k}$ 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 $Y=1$

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

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

\[ \ell(\beta) = \prod_{i=1}^{n} P(Y_i = y_i | x_i) = \prod_{i=1}^{n} \pi^{y_i} (1 - \pi)^{1-y_i} \]

\[ = \prod_{i=1}^{n} \left( \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}} \right)^{y_i} \left( 1 - \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}} \right)^{1-y_i} \]

\[ = \prod_{i=1}^{n} \left( \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}} \right)^{y_i} \left( \frac{1}{1 + e^{x_i' \beta}} \right)^{1-y_i} \]

\[ = \prod_{i=1}^{n} e^{y_i x_i' \beta} \]

\[ = \frac{e^{\sum_{i=1}^{n} y_i x_i' \beta}}{\prod_{i=1}^{n} (1 + e^{x_i' \beta})} \]
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 $H_0: \beta_j = 0$.
- 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 will be available from the course website: http://www.utstat.toronto.edu/brunner/oldclass/appliedf12