

DSC 410/510 Multivariate Statistical Methods

Logistic Regression

1

What is Logistic Regression?

- Specialized regression for a binary response
- It is an example of a “generalized linear model”
- Regression coefficients in the variate indicate relative impact of each predictor
- Less affected by violation of normality and homogeneity assumptions than discriminant analysis
- Can accommodate nonmetric predictors by using dummy variables

2

Representing Binary Response

- Logistic regression models the probability of an event occurring
 - ◆ probability bounded by zero and one
 - ◆ follows a logistic (S-shaped) relationship:



- Ordinary (linear) regression is based on normal errors and constant variance
 - ◆ binary responses follow binomial distribution

3

Estimating the Model

- Variate = $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$
- Iterative algorithm finds the “most likely” β -coefficients that fit the logistic model to the data
 - ◆ dependent (response) variable:
 - 1 (if event occurred), 0 (if not)
 - ◆ independent (predictor) variables: x_1, \dots, x_p
- Estimated probabilities = $1/(1+\exp(-\beta_0-\beta_1 x_1-\dots-\beta_p x_p))$
- Odds ratio = $\Pr(\text{event})/(1-\Pr(\text{event})) = \exp(\beta_0+\dots+\beta_p x_p)$
 - ◆ $\exp(\beta_k)$ represents expected change in odds from increasing x_k by one unit, holding other x 's fixed

4

Assessing Fit

- “Most likely” means “-2 log likelihood” is minimized
 - ◆ -2LL provides a measure of fit
 - ◆ compare with “null model”
 - ◆ assess improvement in fit between models
 - ◆ “pseudo R^2 ” is “like” multiple regression R^2
- Classification matrices also indicate “goodness of fit”
- Other more complicated measures of fit are available

5

Selecting Variables

- Significance testing of β -coefficients
 - ◆ Wald statistic leads to a t-test
- Categorical predictors can be incorporated using dummy variable coding
- Stepwise model selection procedures are available
- Residuals can be analyzed to guide model building
- Influence measures can help identify cases that have a strong effect on the results

6

Example

- Illustrative Example
 - ◆ page 314-321
 - ◆ **hatco.xls** data file
 - ◆ “Illustrative Example from p314-321” handout

7