BIST P8110: Applied Regression II

15. Intro to Generalized Linear Models

Qixuan Chen

Department of Biostatistics
Columbia University

Introduction to GLM

- What is generalized linear models (GLM)?
  - a flexible generalization of linear regression that allows for response variables to have distributions other than a normal distribution, such as
    - binomial
    - Poisson
    - negative binomial
- Why we need GLM?
  - Response variables can be various types of data

Model Components in GLM

- The GLM consists of three elements:
  - A probability distribution for the respondent variable
  - A linear predictor $\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$
  - A link function g such that $g\{E(Y|X)\} = \eta$

Linear Regression

- Linear regression is a special case of the GLM
  - $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$
  - $E(Y_i|\mathbf{X_i}) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}$
    - $Y \sim$ Normal distribution
    - $\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$
    - Identity link function $g\{E(Y|X)\} = E(Y|X) = \eta$

Logistic Regression

- Logistic regression is another type of GLM
  - $\text{logit}\{E(Y_i|\mathbf{X_i})\} = \log \frac{Pr(Y_i=1|\mathbf{X}_i)}{Pr(Y_i=0|\mathbf{X}_i)} = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}$
    - $Y \sim$ Bernoulli distribution
    - $\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$
    - Logit link function $g\{E(Y|X)\} = \text{logit}\{E(Y|X)\} = \eta$

Linear versus Logistic Regression

- The fitted values of the linear regression model:

$$\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p.$$

- The fitted values of the logistic regression model:

$$\text{logit}(\widehat{p}) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p,$$

or with some algebra

$$\hat{p} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p}}.$$

Interpretation of β_j

- In linear regression
  - the average change in Y associated with a one-unit increase in X_j, holding all the other covariates constant (adjusting for the other variables in the model)
- In logistic regression
  - the change in log odds of event associated with a one-unit increase in X_j, holding all the other covariates constant
  - or the odds of event is multiplied by e^{β_j} for a one-unit increase in X_j, adjusting for the other variables in the model

Other GLMs

- Other GLMs to be covered include
  - multinomial logistic regression (generalized logit model)
  - ordinal logistic regression (proportional odds model)
  - Poisson regression
  - negative binomial regression