

Analysis of Categorical Data

Multiple Logistic Regression

Multiple Logistic Regression

- k explanatory variables (regressors)
- To model $\pi = P(Y = 1)$

$$\text{logit}(\pi) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

- e^{β_i} is the multiplicative effect on the odds when $X_i \rightarrow X_i + 1$, all other X s fixed.

Example 3: Horseshoe Crabs Again

Building and Applying Logit Models

- Two competing goals: model complex enough to fit the data well, yet simple enough to easily interpret.
- “All models are wrong, some are useful.” G. Box
- Approximately, how many regressors are allowed?
 - ◆ min: 10 outcomes of each type for every regressor
 - ◆ example: if $Y = 1$ only 30 times, then only three regressors
- Be careful of (multi-)collinearity. e.g. weight and width in the crab example.
- Stepwise variable selection
- Akaike information criterion: AIC

$$AIC = -2\{\text{maximized log likelihood} - \text{number of parameters}\}$$

Example: Clinical Trials Example