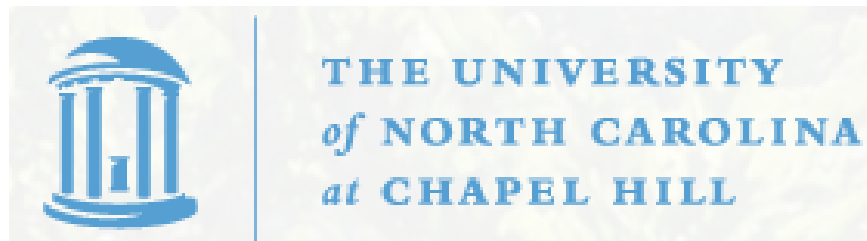


STOR 590:
ADVANCED LINEAR MODELS
Instructor: Richard L. Smith

Class Notes:
August 21, 2020



Overview of Chapter 2

- Initial graphical displays
- Fitting models by logistic regression
- Inference - confidence intervals and tests
- Diagnostics
- Model selection
- Goodness of fit
- Problems with estimation

LOGISTIC REGRESSION

- y_i is 0 or 1, covariates x_{ij} , $0 \leq j \leq p$, $1 \leq i \leq n$.
- Define $p_i = \Pr \{y_i = 1 \mid x_{i0}, \dots, x_{ip}\}$.
- $p_i = \sum_{j=0}^p x_{ij}\beta_j$ makes no sense
- Instead, define $\text{logit}(p) = \log \left(\frac{p}{1-p} \right)$.
- $\text{logit}(p_i) = \sum_{j=0}^p x_{ij}\beta_j$ or $p_i = \frac{\exp(\sum_{j=0}^p x_{ij}\beta_j)}{1 + \exp(\sum_{j=0}^p x_{ij}\beta_j)}$.
- Fit in R by a command of form
`glmod=glm(y~x1+x2,family=binomial)`
with any number of covariates in the sum.

METHOD OF MAXIMUM LIKELIHOOD

- Y_1, \dots, Y_n are observations.
- Density of Y_i is $f_i(\cdot ; \theta)$ where θ is a vector of parameters
 - Density may refer to discrete case (probability mass function), continuous case (pdf) or a mixture of discrete and continuous (e.g. thresholded or censored data)
- Likelihood function $L(\theta) = \prod_{i=1}^n f_i(Y_i ; \theta)$.
- Maximum likelihood estimator (MLE) chooses $\hat{\theta}$ to maximize $L(\theta)$ or equivalently to minimize $\ell(\theta) = -\sum_{i=1}^n \log f_i(Y_i ; \theta)$.

Variances, Covariances, Standard Errors

- Notation: $\frac{\partial^2 \ell}{\partial \theta \partial \theta^T}$ matrix of second-order derivatives ((i, j) entry is $\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}$).
- Let $H(\theta)$ be $\frac{\partial^2 \ell}{\partial \theta \partial \theta^T}$ (Hessian matrix) and let $I(\theta)$ be the expected value of $H(\theta)$
- Usually, $H(\theta)$ is evaluated at the MLE $\hat{\theta}$ and $I(\theta)$ is evaluated at the true value, say θ^* . Then I is the *Fisher Information Matrix* and H is the *Observed Information Matrix*
- Either of the inverses, I^{-1} or H^{-1} is a good approximation to the variance-covariance matrix of $\hat{\theta}$ but H^{-1} is easier to compute
- The square roots of the diagonal entries of H^{-1} are the (estimated) *standard errors* of the parameter estimates
- Aside: No connection with the hat matrix