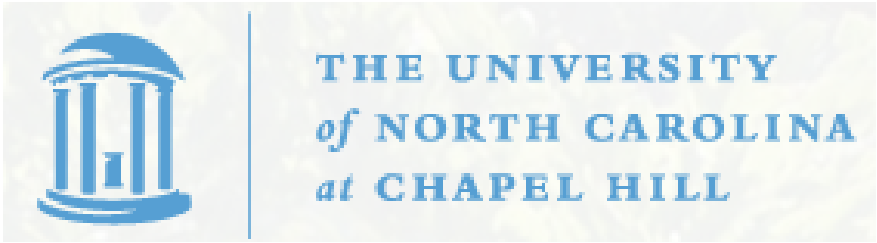
STOR 556: ADV METH DATA ANAL
Instructor: Richard L. Smith

Class Notes #10:

February 12, 2019



Scheduling a Take-home Midterm/Final

- Midterm, posted noon Feb 24, email solutions no later than 6pm Feb 25
- Final, posted noon Apr 30, email solutions no later than 6pm May 1
- Dates are confirmed but will I work with any individual students who have difficulties with those dates

Homework 4

- Chapter 3, Problems 1 and 3
- Hint for problem 1: you can test for interactions by including terms like

  glm(cbind(ncases,ncontrols)~agegp+alcgp+tobgp+agegp*alcgp +agegp*tobgp+alcgp*tobgp,family=binomial,esoph)

  The * terms denote interactions between factor variables.

  Part (c) is open-ended: try to find some model that fits better than the best model from (b)
- Problem 3: data(seeds)
- In both problems, also answer part (i): would the fit be improved by using a quasi-binomial model?
- Due date: Tuesday, February 19.

Summary of Last Class

- Data structure: response is a two-column matrix representing y_i and $m_i - y_i$ in a Binomial(m_i, p_i) experiment
- $\mathrm{E}\{y_i\} = m_i p_i$, $\mathrm{Var}\{y_i\} = m_i p_i(1 - p_i)$
- Logit link: $\eta_i = \log \frac{p_i}{1-p_i} = \sum_{j=0}^{q} x_{ij}\beta_j$
- Fit with `glm` command with `family=binomial`, same as Ch. 2
- Extension: sometimes better to write $\mathrm{Var}\{y_i\} = \phi m_i p_i(1-p_i)$ where ϕ is called the *overdispersion parameter* (typically, but not necessarily, $\phi > 1$)
- This may be fitted using `family=quasibinomial`

Data on Proportions

- Sometimes data consist of proportions without any information about sample sizes
- Here "quasibinomial" is again an option, but we may find overdispersion is < 1
- Alternative: beta model (Section 3.6)

Estimating Using the Beta Distribution

- Density $f(y\ ;\ a,b)\ =\ \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}y^{a-1}(1-y)^{b-1},\ 0<y<1.$
- Mean μ, variance $\frac{\mu(1-\mu)}{(1+\phi)}$, $\mu\ =\ \frac{a}{a+b},\ \phi\ =\ a+b.$
- $\eta\ =\ \log\frac{\mu}{1-\mu}$ is *link function*
- $\eta_i\ =\ \sum_{j=0}^{q}\beta_j x_{ij}$
- Implement with `gam(...,family=betar())` function in package `mgcv`.