4/13/2010 36-402/608 ADA-II H. Seltman

Handout #22: Poisson Regression

1. Context: Poisson outcome with any explanatory variables (defines the error model)
2. The Poisson distribution
$$\Pr(Y = y|\mu) = e^{-\mu}\mu^y/y!$$
has integer outcomes in $[0, \infty]$, so is a good model for event counts. It can arise from a situation where the chance of an event is constant over any fixed amount time (or space), proportional to the available time (or space), and when the chance of two events in a very small time (or space) is negligible. The variance is equal to the mean.
3. Means model: log of the expected count = $\log(\mu_i|X_i) = \mathbf{X_i}\beta$. (Called a "log link function".)
4. Estimates: maximum likelihood via generalized linear model (iterative)
5. Assumptions: independence, single event rate each X combination, linearity on the log scale
6. EDA: Plot log Y vs. X.
7. Residual vs. fit and residual vs. x plots are useful for detecting non-linearity or variance unequal to the mean (under/over dispersion).
8. Likelihood ratio test: $-2\log(\text{Likelihood}_R/\text{Likelihood}_F) \sim \chi^2(\Delta d)$
Alternate form: $\text{Deviance}_R - \text{Deviance}_F \sim \chi^2(\Delta d)$
9. Model selection: LRT, BIC, AIC, forward/backward selection
10. If different units have different exposures, a model "offset" must be used.
11. Coefficient interpretation (can change "associate" to "cause" in appropriate randomized experiments):
    (a) β_0 is the log of the event rate when all explanatory variables are zero
    (b) β_j is the difference in log of the event rate when X_j goes up by 1 unit. $\exp(\beta_j)$ is the corresponding rate ratio.
    (c) *Mutatis mutandis* for indicators and interactions.