Poisson Regression

Some slides from Craig Burkett
Elephants example

- 41 male elephants from Kenyan National Park
  - Followed for 8 years
- Interested in relationship between age & number of successful matings
- (Elephants, in R)
- Can’t use linear regression because responses are not normal
  - Unless matings happen *really* often

![Poisson distribution chart](image)
Elephants Example - Binomial

• Can’t use binomial response because the number of trials isn’t fixed
  • What does a fixed number of trials entail?
Generalized Linear Models -- Review

\[ g(\mu) = X\beta \]

\[ Y \sim \text{dist}_\mu. Y^{(i)} \sim \text{dist}_\mu(g^{-1}(X^{(i)}\beta)) \]

\( g(\mu) \) is called a **Link Function**. The distribution of the Y’s is the family

\( g(\mu) = \mu \) is linear regression – Identity Link

\( g(\mu) = \log(\mu) \) is Log Link – the default for Poisson regression

\( g(\mu) = \log\left(\frac{\mu}{1-\mu}\right) \) is Logistic regression – Logit Link
Estimating GLMs

• Find a $\beta$ such that

$$\Pi_i dist_\mu (g^{-1}(X^{(i)} \beta))$$

is maximized
Response

- Counts, so model with Poisson distribution
  \[ P(Y = y) = \frac{\mu^y e^{-\mu}}{y!} \]
  \[ E[Y] = Var[Y] = \mu \]
- Poisson regression model
  \[ \log(\mu) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \]
Interpretation

• $\mu = \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)$

• Increasing $x_j$ by one unit with other predictors held constant, increases mean of response by factor $e^{\beta_j}$

• Inference (i.e., CI’s for the $\beta$s)
  • Wald and LRT as with logistic
A representation of a log-linear model in which the distribution of $Y$ (as a function of $X$) is Poisson with mean $\mu$ and $\log(\mu) = -1.7 + 0.20X$; the histograms are the Poisson distribution at three values of $X$. 

Visualizing the SD
Conclusions

• Does the average number of matings increase with age?
  • Yes

• For every additional year of age, the expected number of matings increases by
  • $\exp(0.0687) = 1.07$