Poisson Regression

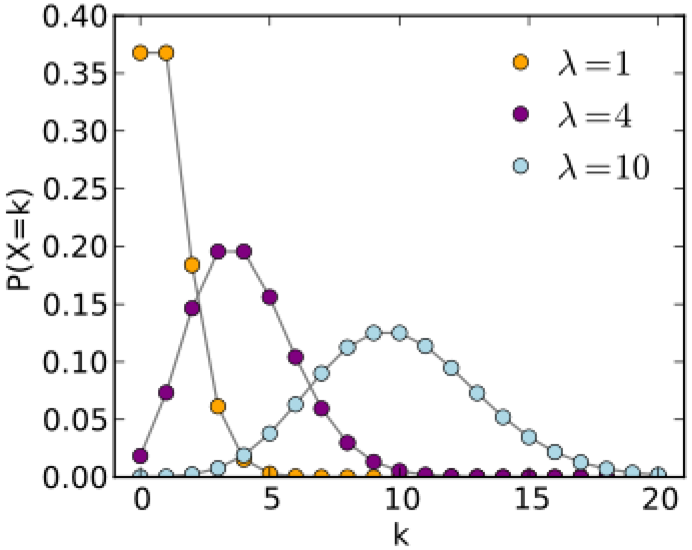
Some slides from Craig Burkett

STA303/STA1002: Methods of Data Analysis II, Summer 2016

Michael Guerzhoy

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

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

$$g(\mu) = X\beta$$
$$Y \sim dist_\mu . Y^{(i)} \sim dist_\mu(g^{-1}(X^{(i)}\beta))$$

$g(\mu)$ 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 β such that
$$\Pi_i dist_\mu(g^{-1}(X^{(i)}\beta))$$
  is maximized

Response

- Counts, so model with Poisson distribution

$$P(Y = y) = \frac{\mu^y e^{-\mu}}{y!}$$

$$E[Y] = Var[Y] = \mu$$

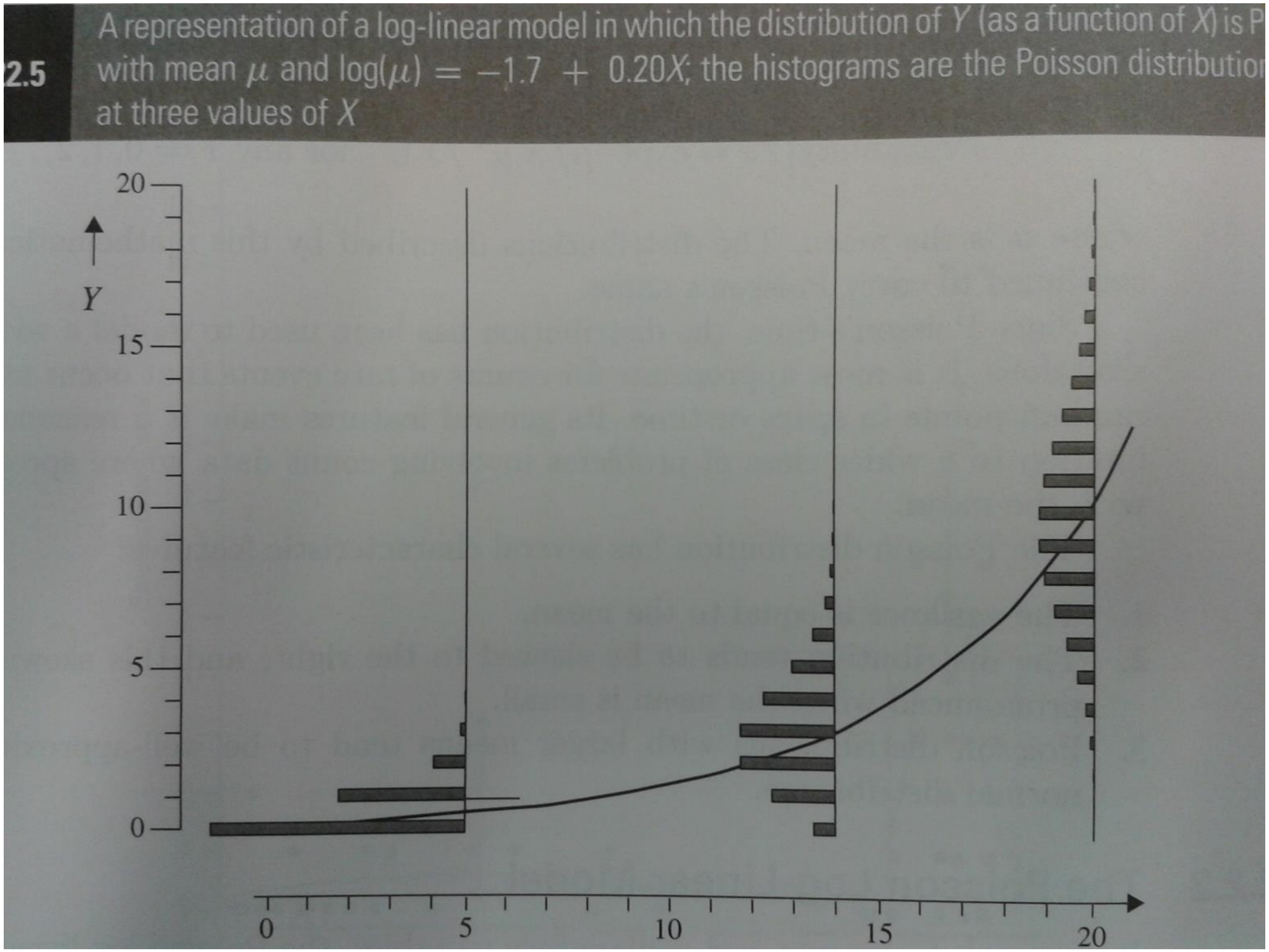
- Poisson regression model

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

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^{β_j}
- Inference (i.e, CI's for the βs)
  - Wald and LRT as with logistic

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