

# Generalized Linear Models Case Studies

Bret Larget

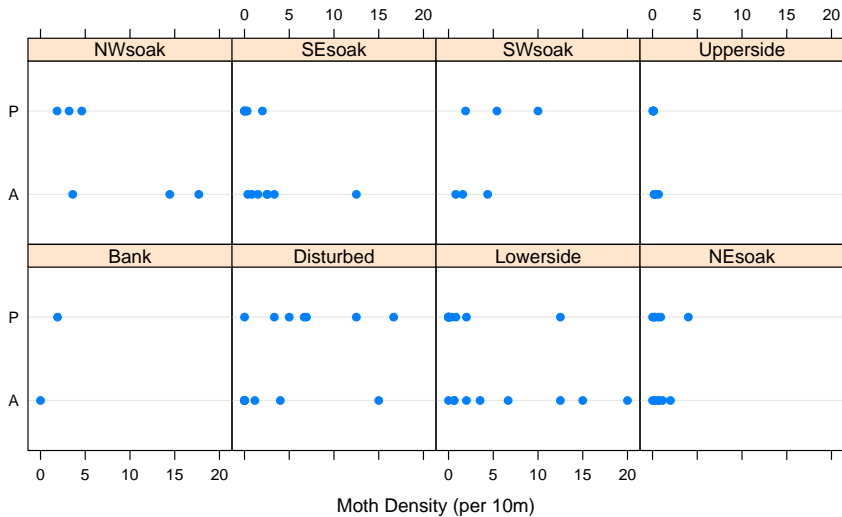
Departments of Botany and of Statistics  
University of Wisconsin—Madison

February 26, 2008

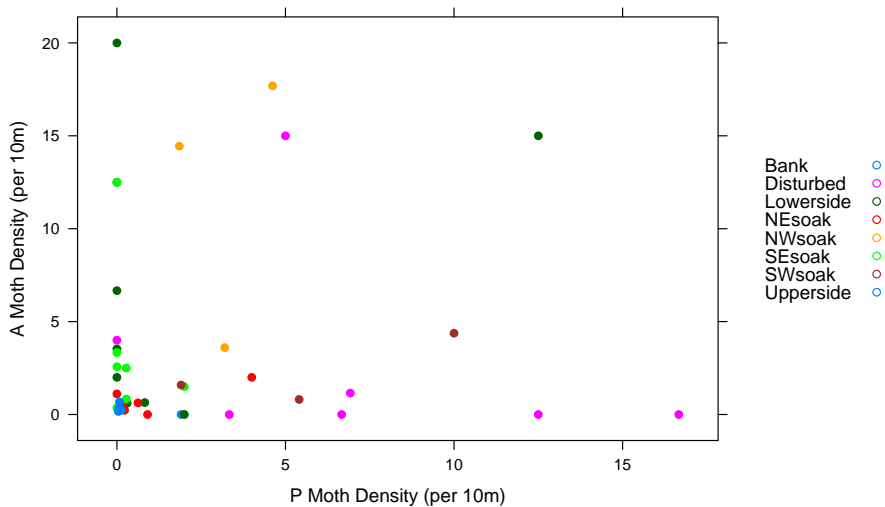
# Poisson Regression Moth Example

- Researchers studied the effect of habitat on density of two different moth species.
- Researchers looked for moths along transects.
- Each transect was partitioned into different sections depending on the habitat type.
- There are 8 different types of habitat.
- Counts of two moth species (labeled A and P) were obtained for each part of each transect.
- Under a Poisson model, we expect the counts to be proportional to the length of the transect within each habitat.

# Dotplot of Data



# Scatterplot of Data



## Model with an Offset

- The count data for species A comes from transects of different lengths.
- Poisson process assumptions suggest that the expected count for a single transect *should be proportional to the length* of the transect.
- We need to modify our typical Poisson regression model to accommodate an *exposure* variable.

- For transect  $i$ :
  - ▶  $u_i$  is the length in meters;
  - ▶  $h_i$  is the habitat;
  - ▶  $\theta_i = \theta(h_i)$  is the rate;
  - ▶  $y_i$  be the count of species A;
- The expected count for the  $i$ th habitat is then:

$$E[y_i] = \mu_i = u_i\theta_i$$

## Derivation (cont.)

- We model  $\theta_i = \exp(X_i\beta)$  where  $X_i$  includes all predictors *except for length*.
- In this case:
  - ▶  $\theta_i = \exp(\beta_1)$  if  $h_i$  is the first habitat;
  - ▶  $\theta_i = \exp(\beta_1 + \beta_k)$  if  $h_i$  is the  $k$ th habitat,  $k > 1$ .
- We need  $E[y_i] = \mu_i = u_i\theta_i$ .
- With the exponential inverse link function, this is:

$$\begin{aligned}\mu_i &= \exp(\log u_i + \log(\theta_i)) \\ &= \begin{cases} \exp(\log u_i + \beta_1) & \text{if } k = 1 \\ \exp(\log u_i + \beta_1 + \beta_k) & \text{if } k > 1 \end{cases}\end{aligned}$$

- This is like regular Poisson regression with  $\log u$  as a predictor, *except we want to force the estimated coefficient to be equal to one*.

## Fitting the Model

```
> fitA = with(moths, glm(A ~ habitat, data = moths, family = poisson,  
+   offset = log(length)))  
> fitAq = with(moths, glm(A ~ habitat, data = moths, family = quasipoisson,  
+   offset = log(length)))  
> fitP = with(moths, glm(P ~ habitat, data = moths, family = poisson,  
+   offset = log(length)))  
> fitPq = with(moths, glm(P ~ habitat, data = moths, family = quasipoisson,  
+   offset = log(length)))
```

# Poisson Regression for A

```
> display(fitA)
```

```
glm(formula = A ~ habitat, family = poisson, data = moths, offset = log(len
```

	coef.est	coef.se
(Intercept)	-18.35	1275.75
habitatDisturbed	16.53	1275.75
habitatLowerside	16.81	1275.75
habitatNEsoak	15.45	1275.75
habitatNWsoak	18.44	1275.75
habitatSEsoak	16.70	1275.75
habitatSWsoak	16.59	1275.75
habitatUpperside	14.82	1275.75

```
---
```

```
n = 41, k = 8
```

```
residual deviance = 180.3, null deviance = 499.0 (difference = 318.7)
```

# QuasiPoisson Regression for A

```
> display(fitAq)
```

```
glm(formula = A ~ habitat, family = quasipoisson, data = moths,  
     offset = log(length))
```

	coef.est	coef.se
(Intercept)	-18.35	3793.81
habitatDisturbed	16.53	3793.81
habitatLowerside	16.81	3793.81
habitatNEsoak	15.45	3793.81
habitatNWsoak	18.44	3793.81
habitatSEsoak	16.70	3793.81
habitatSWsoak	16.59	3793.81
habitatUpperside	14.82	3793.81

```
---
```

```
n = 41, k = 8
```

```
residual deviance = 180.3, null deviance = 499.0 (difference = 318.7)
```

```
overdispersion parameter = 8.8
```

# Poisson Regression for P

```
> display(fitP)
```

```
glm(formula = P ~ habitat, family = poisson, data = moths, offset = log(len
```

```
      coef.est coef.se
(Intercept)  -1.66    0.50
habitatDisturbed  1.26    0.53
habitatLowerside -0.76    0.56
habitatNEsoak   -1.24    0.57
habitatNWsoak    0.43    0.55
habitatSEsoak   -1.81    0.65
habitatSWsoak    0.78    0.52
habitatUpperside -3.12    0.61
```

```
---
```

```
n = 41, k = 8
```

```
residual deviance = 92.6, null deviance = 389.5 (difference = 296.9)
```

# QuasiPoisson Regression for P

```
> display(fitPq)
```

```
glm(formula = P ~ habitat, family = quasipoisson, data = moths,  
     offset = log(length))
```

	coef.est	coef.se
(Intercept)	-1.66	1.07
habitatDisturbed	1.26	1.13
habitatLowerside	-0.76	1.19
habitatNEsoak	-1.24	1.21
habitatNWsoak	0.43	1.18
habitatSEsoak	-1.81	1.38
habitatSWsoak	0.78	1.11
habitatUpperside	-3.12	1.31

```
---
```

```
n = 41, k = 8
```

```
residual deviance = 92.6, null deviance = 389.5 (difference = 296.9)
```

```
overdispersion parameter = 4.6
```

- More live with R and on Board