

# Poisson Regression

Bret Larget

Departments of Botany and of Statistics  
University of Wisconsin—Madison

May 3, 2007

## Moth Example

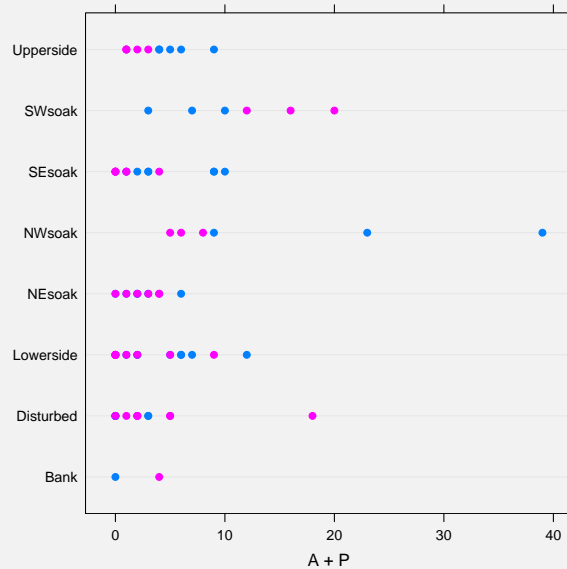
### Poisson Regression Moth Example

- Researchers studied the effect of habitat on density of two different moth species.
- Over a period of 16 weeks, researchers counted moths at 20 different times.
- Researchers looked for moths along transects of widely varying lengths (from 2 to 233 meters).
- Each transect was partitioned into different sections depending on the habitat type.
- There are 8 different habitat types.
- Counts of two moth species (labeled A and P) were obtained for each part of each transect.

## Dotplot of Data

```
> library(lattice)
> library(DAAG)
> dotplot(habitat ~ A + P, data = moths, pch = 16)
```

- A is in blue
- P is in purple



## Table of Data

```
> attach(moths)
> Number = table(habitat)
> Transect = round(sapply(split(meters, habitat), mean), 1)
> MothA = sapply(split(A, habitat), sum)
> MothP = sapply(split(P, habitat), sum)
> rbind(Number, Transect, MothA, MothP)
```

	Bank	Disturbed	Lowerside	NEsoak	NWsoak	SEsoak	SWsoak	Upperside
Number	1	7	9.0	6.0	3.0	7.0	3.0	5.0
Transect	21	7	21.2	42.3	21.7	27.6	38.7	190.4
MothA	0	8	41.0	14.0	71.0	37.0	20.0	28.0
MothP	4	33	17.0	14.0	19.0	6.0	48.0	8.0

## Analysis of Species A

- Since transect lengths (in meters) are on such different scales, we will log transform them.
- Using only habitat as an explanatory variable is similar to one-way ANOVA — each habitat type gets its own mean.
- Using both categorical habitat and continuous  $\log(\text{transect length})$  is similar to Analysis of Covariance — we fit separate curves for each habitat type.

## Analysis of Species A

- Mean 0 for Bank habitat will cause trouble as the estimated parameter value would be  $-\infty$ .
- Analyze data without this observation.
- No obvious control group, so change the default parameterization.
- Use quasi-Poisson model (anticipating more spread than Poisson model).

```
> noBank = (habitat != "Bank")
> form.a = A ~ habitat + log(meters)
> options(contrasts = c("contr.sum", "contr.poly"))
> a.glm = glm(form.a, family = quasipoisson, subset = noBank)
```

## Summary of Analysis of Species A

```
> summary(a.glm)

Call:
glm(formula = form.a, family = quasipoisson, subset = noBank)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.55619  -1.47034  -0.04348   0.93818   3.08215

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.15105    0.51257   2.246  0.0318 *
habitat1    -1.22447    0.56484  -2.168  0.0377 *
habitat2     0.05876    0.28286   0.208  0.8367
habitat3    -0.76294    0.39992  -1.908  0.0654 .
habitat4     1.62119    0.21720   7.464 1.7e-08 ***
habitat5     0.12083    0.26778   0.451  0.6549
habitat6     0.29020    0.34351   0.845  0.4045
log(meters)  0.12917    0.15278   0.845  0.4041
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 2.700832)

Null deviance: 246.293  on 39  degrees of freedom
Residual deviance: 93.991  on 32  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

## Comments

- The estimated dispersion parameter is 2.7, substantially larger than 1.
- There is little evidence that the transect length makes much difference.
- (This is explained by the observation in the field the preferred habitat is near the start of each transect.)
- The 4th habitat (not counting Bank) is NWsoak which has the largest mean.
- Two of these habitats have mild evidence that the means are smaller than is typical.

## Estimated Model

- The estimated model is as follows:

$$\begin{aligned} E(\text{count of } A) &= \exp\{1.15 + 0.129 \log(\text{meter}) + \text{habitat effect}\} \\ &= 3.16 \times (\text{meter})^{0.129} \times \exp\{\text{habitat effect}\} \end{aligned}$$

- The habitat effect parameters are:

Disturbed	Lowerside	NEsoak	NWsoak	SEsoak	SWsoak	Upperside
-1.22	0.06	-0.76	1.62	0.12	0.29	-0.10

- For example, the predicted count for a Disturbed transect of length 3m would be

$$3.16 \times (3)^{0.129} \times \exp(-1.22) = 1.071$$

## Analysis of P

- If time permits, do analysis for species P.