Generalized Linear Models

The general linear model

- Forms the basis for most classical statistics

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon_i$$

$$= \beta \boldsymbol{X} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2 I)$$

The Generalized Linear Model

Strategy - "Generalize" the general linear model to allow for

- non-normal distributions of errors
- certain non-linear relationships

$$Y_i = \beta \boldsymbol{X} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2 I)$$

$$Y_i = f(\beta \boldsymbol{X}) + \varepsilon_i \qquad \varepsilon_i \sim \text{other pdf's}$$

$$f^{-1}(Y_i) = \beta \boldsymbol{X}$$

Generalized Linear Model (GLM, or GLiM)

(Not to be confused with SAS's proc GLM)

The Generalized Linear Model

$$f^{-1}(Y_i) = \beta \boldsymbol{X} \qquad \varepsilon_i \sim \text{other pdf's}$$

"Link" Function

- Maximum Likelihood Estimation of parameters
- Measure of fit is the "Deviance"

The Generalized Linear Model

- For example
  - Count data

$$\ln(Y_i) = \beta_0 + \beta_1 X \qquad \varepsilon_i \sim \text{Poisson}$$

Log-linear Regression

The Generalized Linear Model

- For example
  - Proportion data

Let response variable be p_i

Logit(p)

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X \qquad \varepsilon_i \sim \text{Binomial}$$

Logit Regression = Logistic Regression

The Generalized Linear Model

- For example
  - Proportion data

Let response variable be p_i

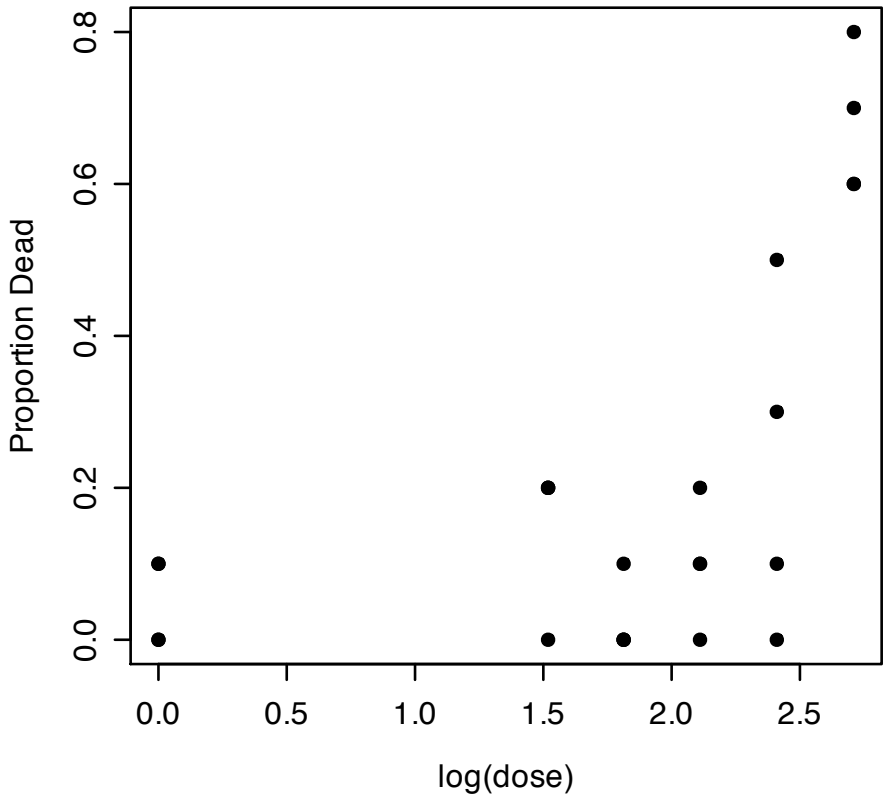
$$\text{Inverse Normal CDF}(p_i) = \beta_0 + \beta_1 X \qquad \varepsilon_i \sim \text{Binomial}$$

probit(p)

Probit Regression

The Generalized Linear Model

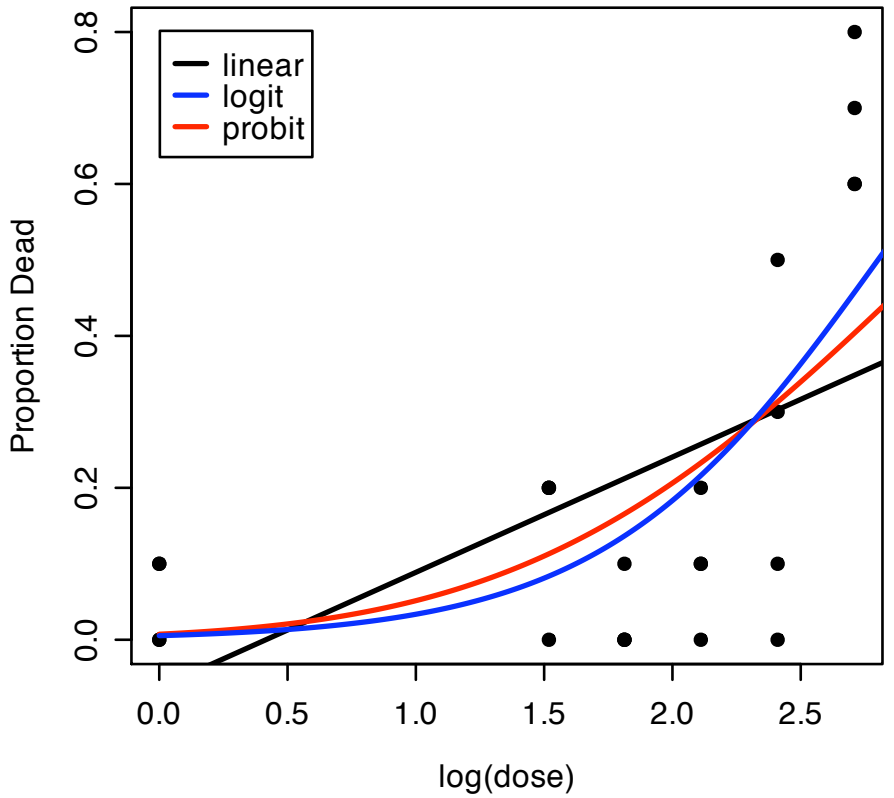
Fathead minnow (*Pimephales promelas*) larvae were exposed to NaPCP. Ten fish were exposed in each tank. Four tanks per dose.



From Weber et al. (1989) (1991) as described in Newman, M. C.1995. *Quantitative Methods in Aquatic Toxicology*. CRC Press, Boca Raton, FL.

The Generalized Linear Model

Fathead minnow (*Pimephales promelas*) larvae were exposed to NaPCP. Ten fish were exposed in each tank. Four tanks per dose.

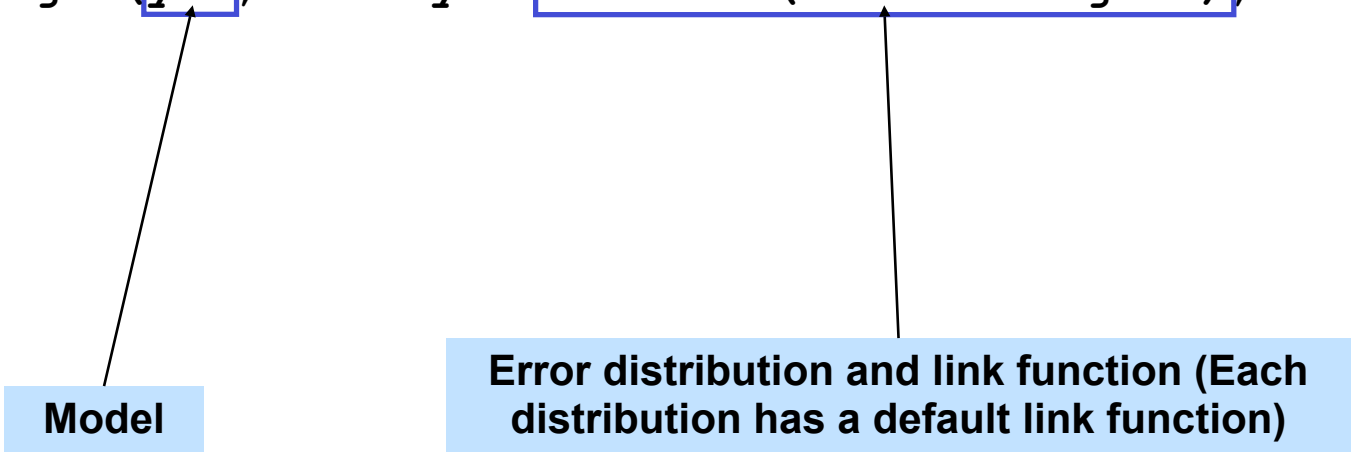


From Weber et al. (1989) (1991) as described in Newman, M. C.1995. *Quantitative Methods in Aquatic Toxicology*. CRC Press, Boca Raton, FL.

The Generalized Linear Model

R implements a procedure called glm

```
> out = glm(y~x, family = binomial(link = "logit"), data)
```



Default Link Functions

```
binomial(link = "logit")
poisson(link = "log")
gaussian(link = "identity")
```

The Generalized Linear Model

R implements a stepwise procedure called glm

```
> out = glm(y~x, family = binomial(link = "logit"), data)
> summary(out)
> anova(out, test = "Chisq")
```

The comparison of Binomial and Poisson glm models is based on a liklihood ratio which is aproximated by a χ^2 distribution rather than F

The Generalized Linear Model

Tobacco budworms (*Heliothis virescens*) were exposed to the pyrethroid *trans*-cypermethrin. Batches of 20 moths of each sex were exposed for 3 days and the number of dead were recorded.

	Dose					
Sex	**1**	**2**	**4**	**8**	**16**	**32**
M	**1**	**4**	**9**	**13**	**18**	**20**
F	**0**	**2**	**6**	**10**	**12**	**16**

From Collett (1991) as described in Venables, W. N., & B. D. Ripley. 2002. *Modern Applied Statistics with S*. Springer, New York.

- Demo