

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2006, The Johns Hopkins University and Karl W. Broman. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.

Spider mites example

Dose of DDT	No. survived	No. dead
0.0	18	7
0.5	19	6
1.0	12	13
1.5	5	20
2.0	6	19
2.5	2	23
3.0	1	24

Binary vs. continuous outcomes

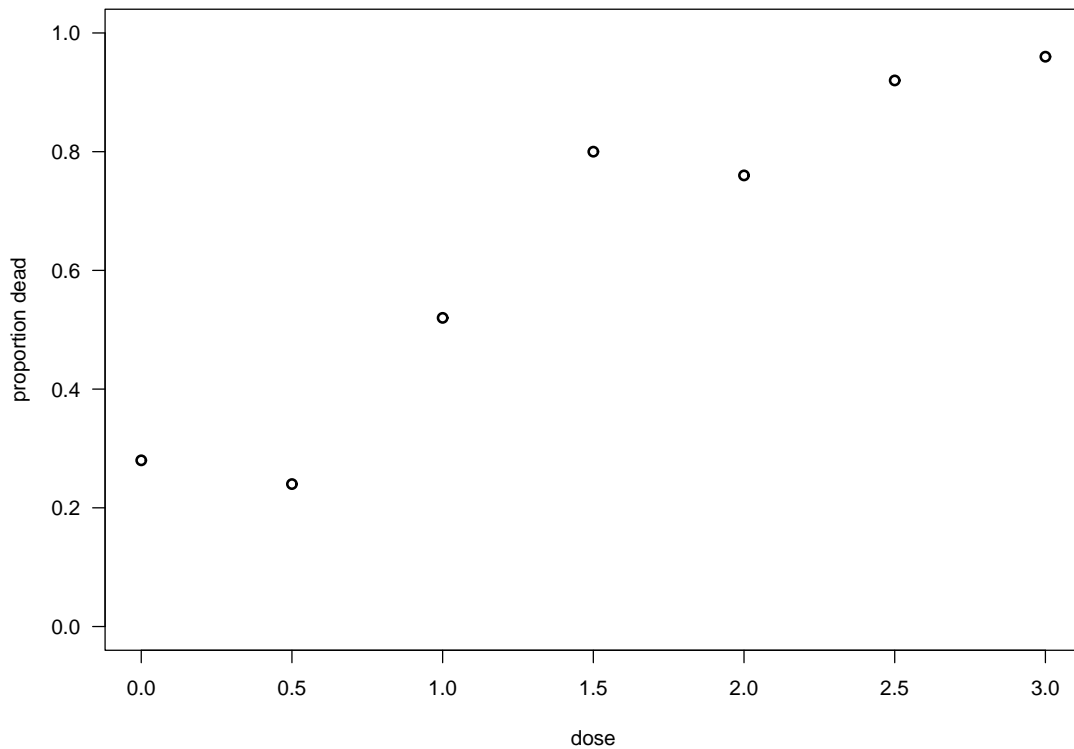
Continuous: ANOVA \longleftrightarrow Regression

Binary: $k \times 2$ table \longleftrightarrow ?

Goals:

- Relationship between dose and $\Pr(\text{dead})$.
- Dose at which $\Pr(\text{dead}) = 1/2$.

A plot of the data



Linear regression

Model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon, \quad \epsilon \sim \text{iid Normal}(0, \sigma^2)$$

This implies:

$$E(y \mid x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

β_i = increase in mean of Y associated with a unit change in x_i

Binary outcomes

Let $p_d = \Pr(\text{dead} \mid \text{dose } d)$

$$p_d = \beta_0 + \beta_1 d ?$$

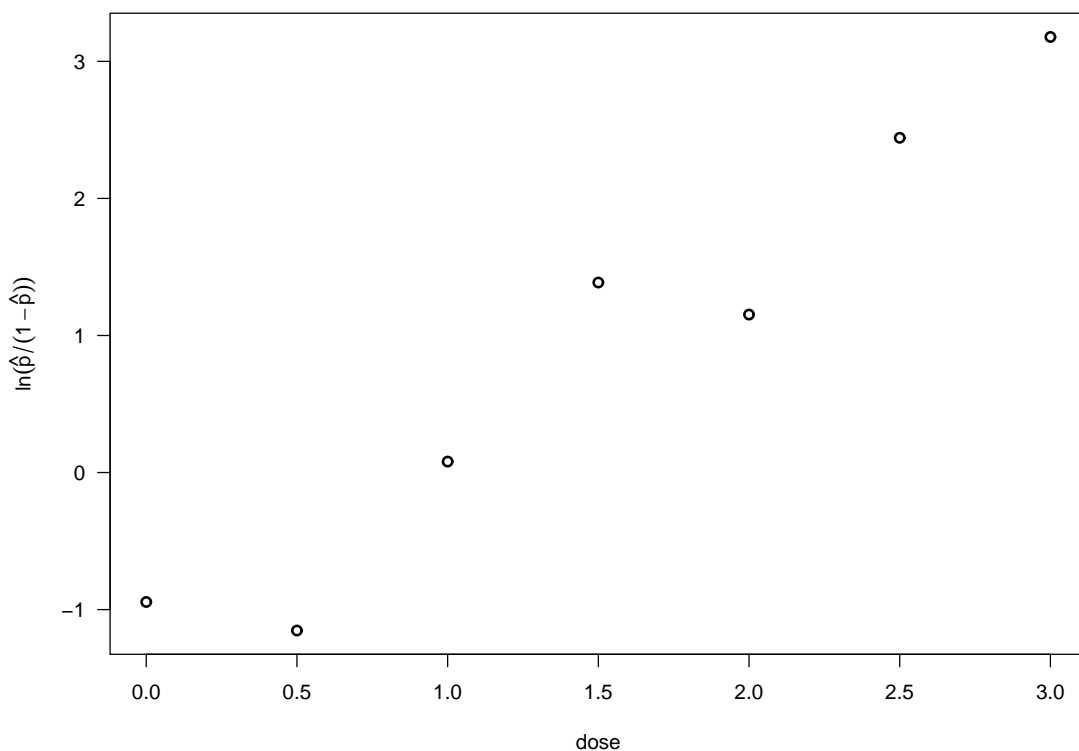
$$0 \leq p_d \leq 1 \quad \text{but} \quad -\infty \leq \beta_0 + \beta_1 d \leq \infty$$

Odds of death: $0 \leq \frac{p_d}{1 - p_d} \leq \infty$

log odds of death: $-\infty \leq \ln\left(\frac{p_d}{1 - p_d}\right) \leq \infty$

$\ln\left(\frac{p}{1 - p}\right)$ is also called $\text{logit}(p)$ or the logistic function.

$\text{logit}(\hat{p}_d)$ vs d



Logistic regression

$$\ln \left(\frac{p_d}{1 - p_d} \right) = \beta_0 + \beta_1 d$$

Try least squares, regressing $\ln \left(\frac{\hat{p}_d}{1 - \hat{p}_d} \right)$ on the dose, d ?

Problems:

- What if $\hat{p}_d = 0$ or 1 ?
- $SD(\hat{p}_d)$ is not constant with d .

Maximum likelihood

Assume $y_d \sim \text{Binomial}(n_d, p_d)$, y_d independent

$$\text{with } \text{logit}(p_d) = \ln \left(\frac{p_d}{1 - p_d} \right) = \beta_0 + \beta_1 d$$

Note:
$$p_d = \frac{e^{\beta_0 + \beta_1 d}}{1 + e^{\beta_0 + \beta_1 d}}$$

Likelihood:

$$L(\beta_0, \beta_1 | \mathbf{y}) = \prod_d p_d^{y_d} (1 - p_d)^{(n_d - y_d)}$$

Logistic regression in R

Logistic regression is a special case of a “generalized linear model”.

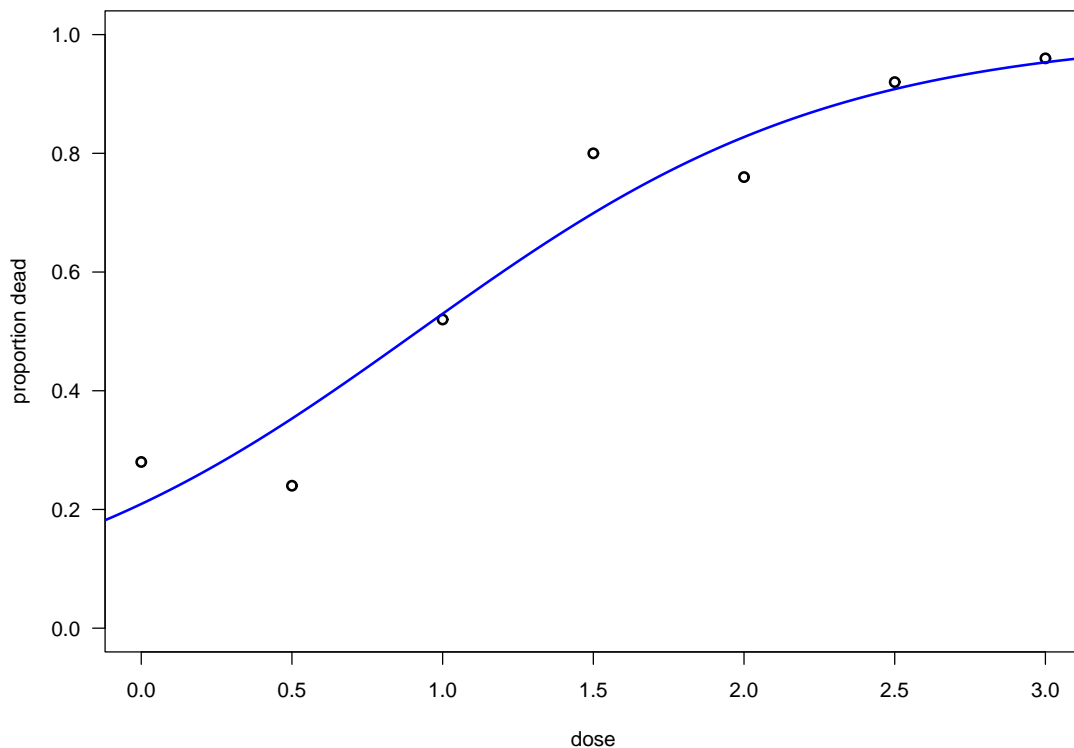
Function in R: `glm()`

```
> glm.out <- glm(n.dead/n ~ dose, weights=n, data=spiders,  
                family=binomial(link=logit))
```

```
> summary(glm.out)$coef
```

	Est	SE	t-val	P-val
(Intercept)	-1.33	0.33	-4.06	<0.001
dose	1.44	0.23	6.29	<0.001

Fitted curve



Interpretation of β 's

$$\ln \left(\frac{p_d}{1 - p_d} \right) = \beta_0 + \beta_1 d$$

β_0 = log odds when dose = 0

Note: $\beta_0 = 0 \longrightarrow p_0 = \frac{1}{2}$

β_1 = change in log odds with unit increase in dose

Note: $\beta_1 = 0 \longrightarrow$ survival unrelated to dose.

LD50

LD50 = dose at which $\Pr(\text{dead} \mid \text{dose}) = \frac{1}{2}$.

$$\ln \left(\frac{1/2}{1 - 1/2} \right) = \beta_0 + \beta_1(\text{LD50})$$

$$\implies 0 = \beta_0 + \beta_1(\text{LD50})$$

$$\implies \text{LD50} = -\beta_0/\beta_1$$

$$\widehat{\text{LD50}} = -\hat{\beta}_0/\hat{\beta}_1$$

$$\widehat{\text{SE}}(\widehat{\text{LD50}}) \approx |\widehat{\text{LD50}}| \sqrt{\left(\frac{\widehat{\text{SE}}(\hat{\beta}_0)}{\hat{\beta}_0} \right)^2 + \left(\frac{\widehat{\text{SE}}(\hat{\beta}_1)}{\hat{\beta}_1} \right)^2 - 2 \frac{\text{cov}(\hat{\beta}_0, \hat{\beta}_1)}{\hat{\beta}_0 \hat{\beta}_1}}$$

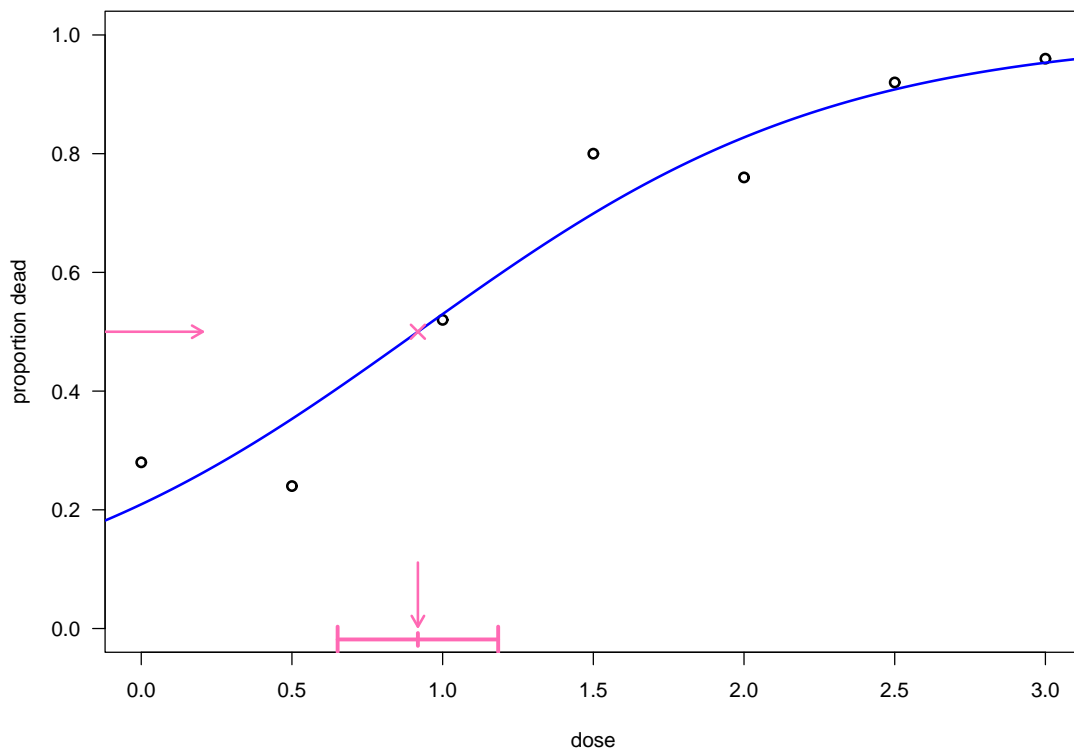
Estimating LD50 in R

```
> glm.out <- glm(n.dead/n ~ dose, weights=n, data=spiders,
                family=binomial(link=logit))
> glm.sum <- summary(glm.out)

> co <- glm.out$coef
> ld50 <- -co[1]/co[2]
> se.co <- glm.sum$coef[,2]
> cov.co <- glm.sum$cov.scaled[1,2]
> se.ld50 <- abs(ld50) * sqrt( (se.co[1]/co[1])^2 +
                              (se.co[2]/co[2])^2 -
                              2*cov.co/(co[1]*co[2]) )

> ld50
  0.92
> se.ld50
  0.14
> ld50 + c(-1,1) * qnorm(0.975) * se.ld50
  0.65  1.18
```

LD50



Another example

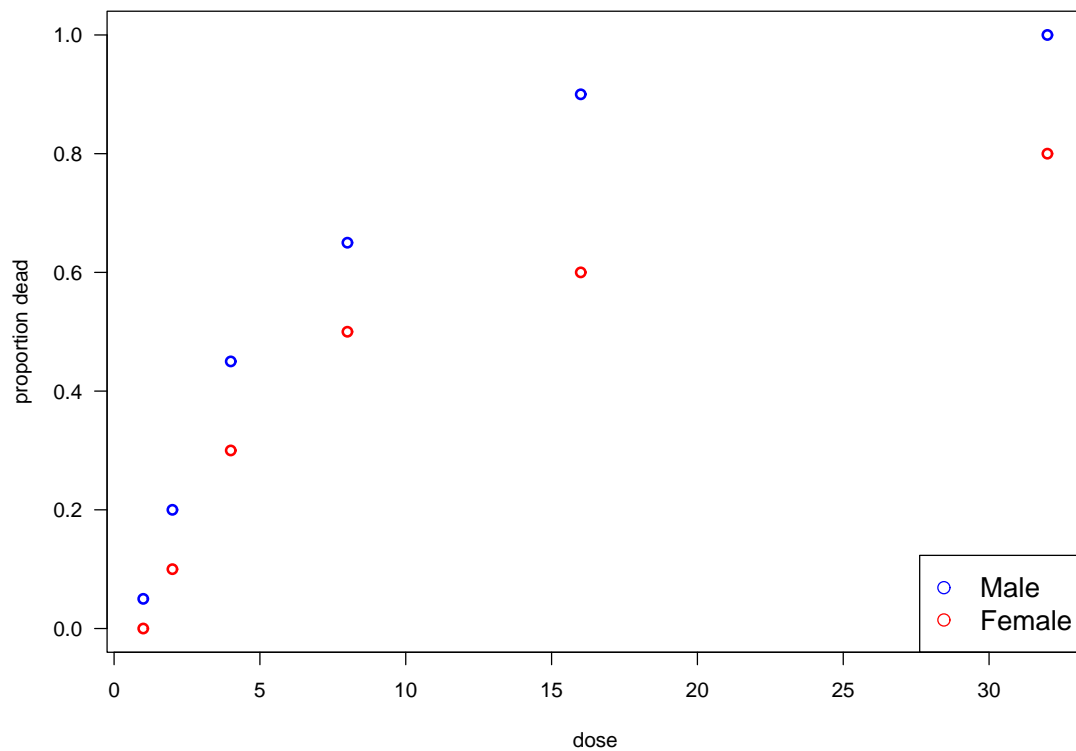
Tobacco budworm, *Heliothis virescens*

Batches of 20 male and 20 female worms were given a 3-day dose of pyrethroid *trans*-cypermethrin

The no. dead or “knocked down” in each batch was noted.

	Dose					
Sex	1	2	4	8	16	32
Male	1	4	9	13	18	20
Female	0	2	6	10	12	16

A plot of the data



Analysis in R

Assume no sex difference

```
> glm.out <- glm(n.dead/n ~ dose, weights=n, data=worms,  
                family=binomial(link=logit))
```

```
> summary(glm.out)$coef  
              Est      SE  t-val  P-val  
(Intercept) -1.57   0.23  -6.8 <0.001  
dose         0.153  0.022   6.8 <0.001
```

Assume sexes completely different

```
> glm.outB <- glm(n.dead/n ~ sex*dose, weights=n, data=worms,  
                 family=binomial(link=logit))
```

```
> summary(glm.outB)$coef  
              Est      SE  t-val  P-val  
(Intercept) -1.72   0.32  -5.3 <0.001  
sexmale      -0.21   0.51  -0.4  0.68  
dose         0.116  0.024   4.9 <0.001  
sexmale:dose 0.182  0.067   2.7  0.007
```

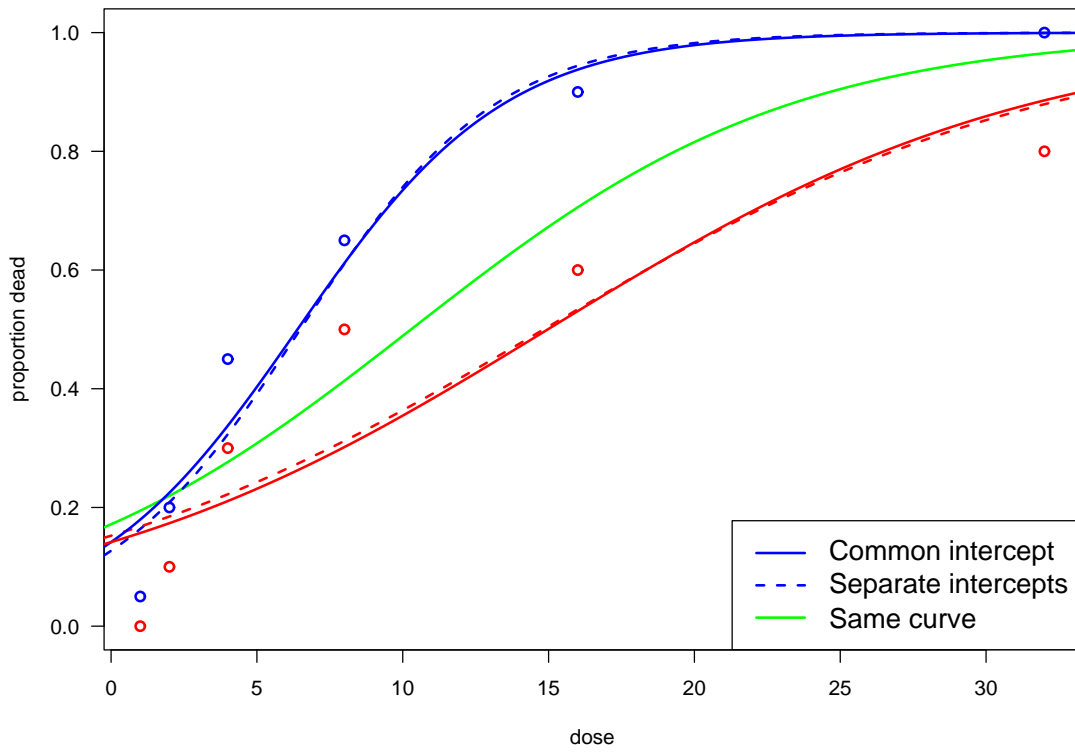
Analysis in R (continued)

Different slopes but common “intercept”

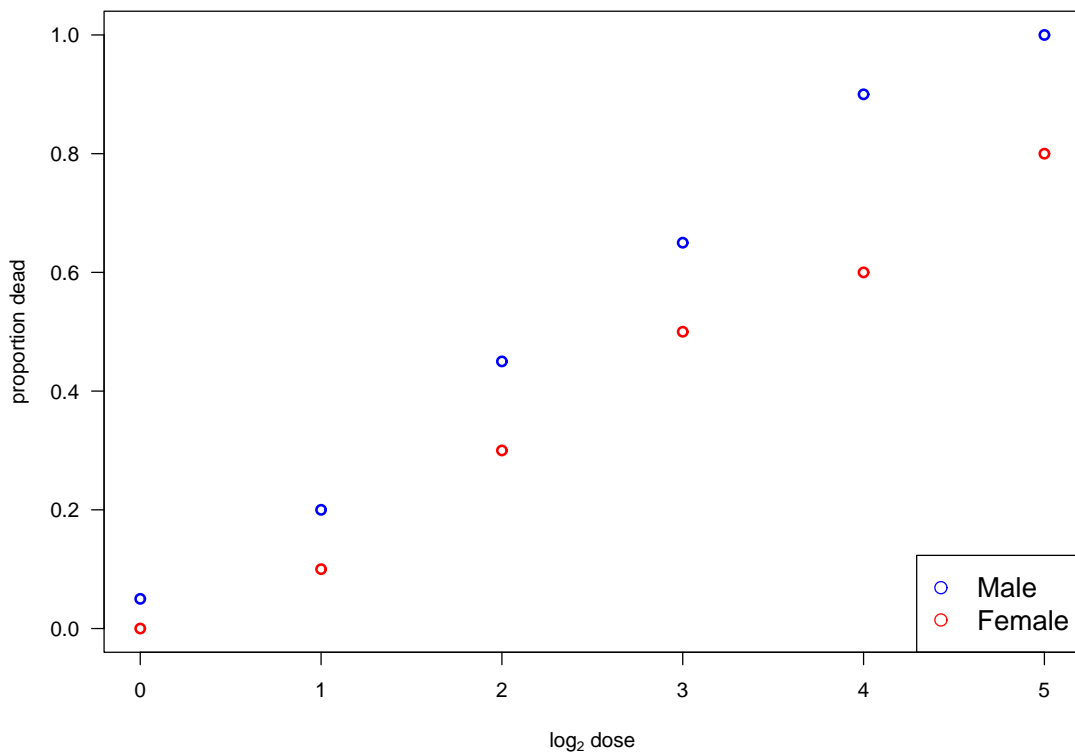
```
> glm.outC <- glm(n.dead/n ~ dose + sex:dose, weights=n,  
                 data=worms, family=binomial(link=logit))
```

```
> summary(glm.out)$coef  
              Est      SE  t-val  P-val  
(Intercept) -1.80   0.25  -7.2 <0.001  
dose         0.120  0.021   5.6 <0.001  
dose:sexmale 0.161  0.044   3.7 <0.001
```

Fitted curves



Plot using \log_2 dose



Use \log_2 of the dose

Assume no sex difference

```
> glm.out <- glm(n.dead/n ~ dose, weights=n, data=worms,  
                family=binomial(link=logit))
```

```
> summary(glm.out)$coef  
              Est      SE  t-val   P-val  
(Intercept) -2.77  0.37   -7.6  <0.001  
dose         1.01  0.12    8.1  <0.001
```

Assume sexes completely different

```
> glm.outB <- glm(n.dead/n ~ sex*dose, weights=n, data=worms,  
                 family=binomial(link=logit))
```

```
> summary(glm.outB)$coef  
              Est      SE  t-val   P-val  
(Intercept) -2.99  0.55   -5.4  <0.001  
sexmale      0.17  0.78   -0.2   0.82  
dose         0.91  0.17    5.4  <0.001  
sexmale:dose 0.35  0.27    1.3   0.19
```

Use \log_2 of the dose (continued)

Different slopes but common “intercept”

```
> glm.outC <- glm(n.dead/n ~ dose + sex:dose, weights=n,  
                 data=worms, family=binomial(link=logit))
```

```
> summary(glm.out)$coef  
              Est      SE  t-val   P-val  
(Intercept) -2.91  0.39   -7.5  <0.001  
dose         0.88  0.13    6.9  <0.001  
dose:sexmale 0.41  0.12    3.3   0.001
```

Fitted curves

