Logistic Regression

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:

$\longrightarrow$ Determine the relationship between dose and Pr(dead).

$\longrightarrow$ Find the dose at which Pr(dead) = 1/2.

A plot of the data

Binary outcomes

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

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

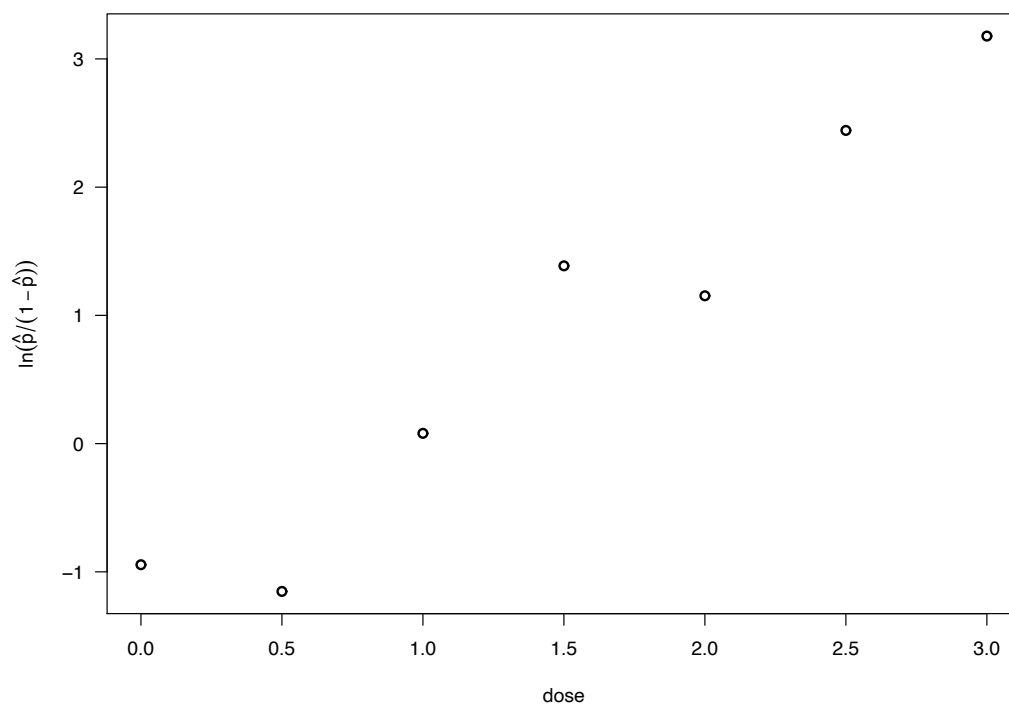$$0 \leq p_d \leq 1 \qquad \text{but} \qquad -\infty \leq \beta_0 + \beta_1 d \leq \infty$$

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

Log odds of death: $\quad -\infty \leq \ln\left(\dfrac{p_d}{1 - p_d}\right) \leq \infty$

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

---

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:

$\longrightarrow$ What if $\hat{p}_d = 0$ or 1?

$\longrightarrow$ $SD(\hat{p}_d)$ is not constant with d.

Maximum likelihood

Assume that

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

$\circ$ y_d independent,

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

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

Likelihood:

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

Logistic regression in R

Logistic regression is a special case of a *generalized linear model*.
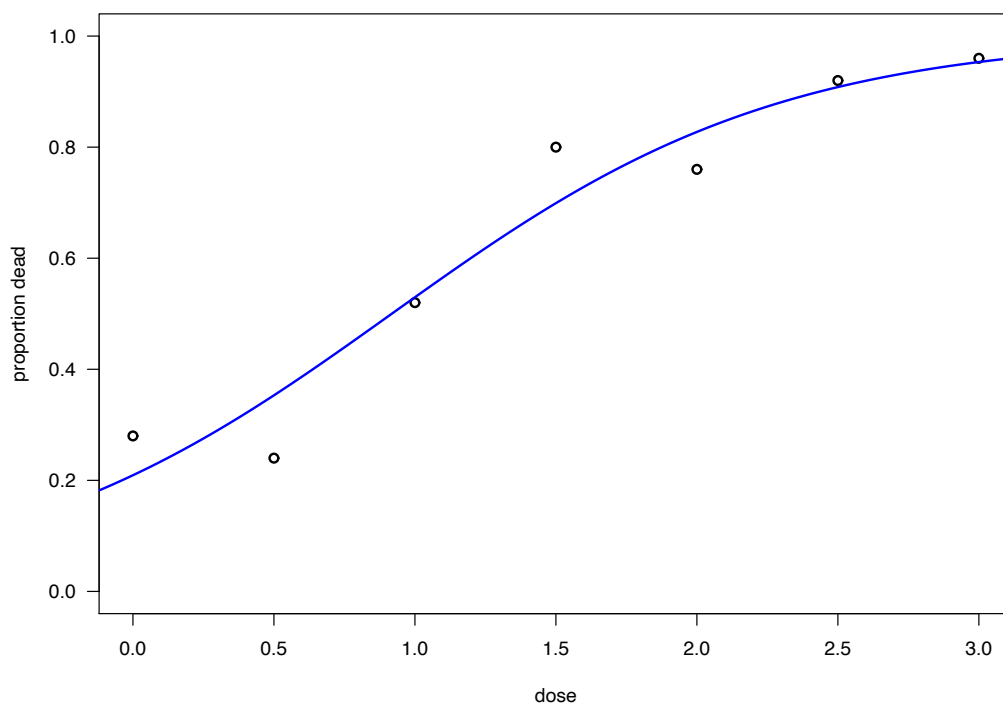
Function in R: `glm()`

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

> summary(glm.out)$coef

                  Est     SE   z-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: $\beta_0 = 0 \quad \longrightarrow \quad p_0 = \frac{1}{2}$

β_1 = change in log odds with unit increase in dose

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

LD50

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

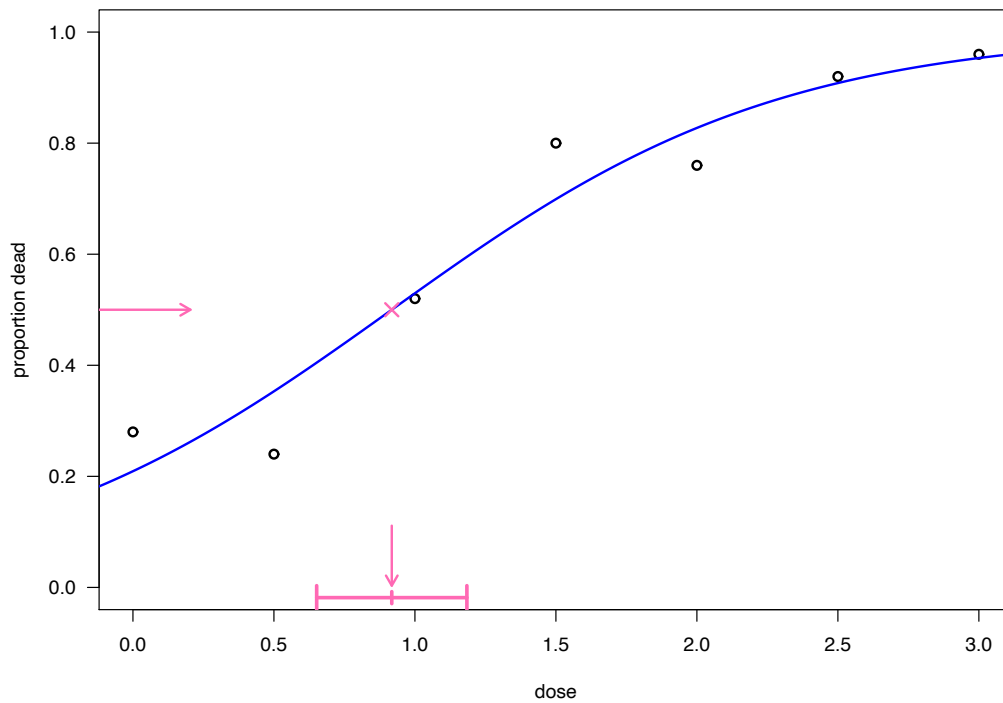$$\ln\left(\frac{1/2}{1 - 1/2}\right) = \beta_0 + \beta_1(\text{LD50})$$

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

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

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

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

LD50

Another example
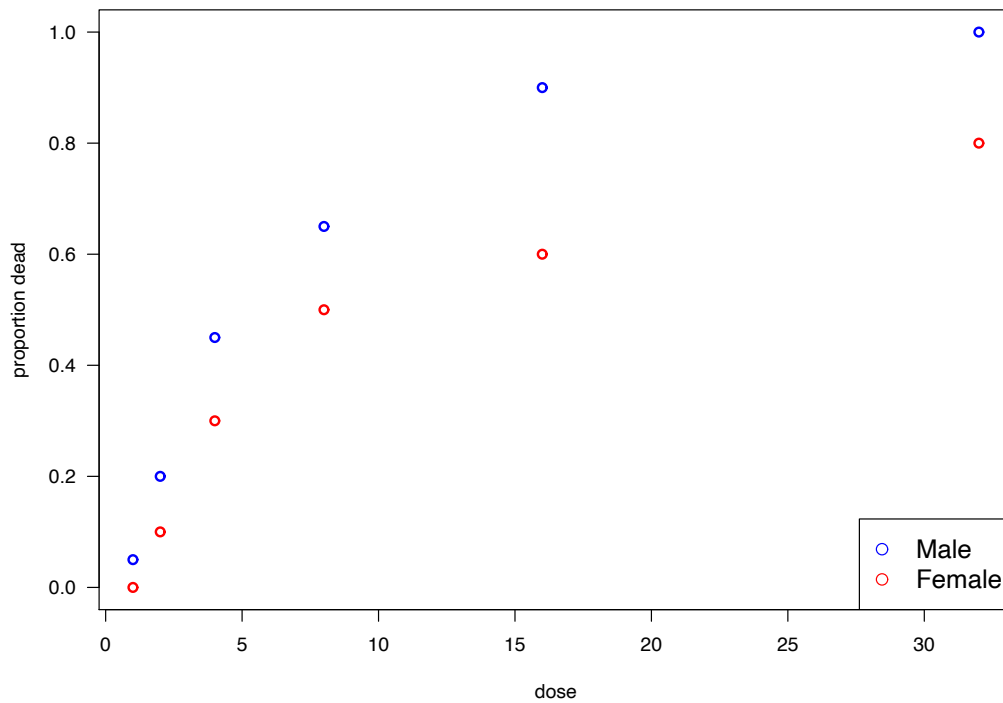
Tobacco budworm, *Heliothis virescens*

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   z-val     P-val
  (Intercept)   -1.57   0.23   -6.81   <0.001
  dose           0.15   0.02    6.81   <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   z-val     P-val
  (Intercept)   -1.72   0.32   -5.32   <0.001
  sexmale       -0.21   0.52   -0.41    0.681
  dose           0.12   0.02    4.86   <0.001
  sexmale:dose   0.18   0.07    2.71    0.007
```
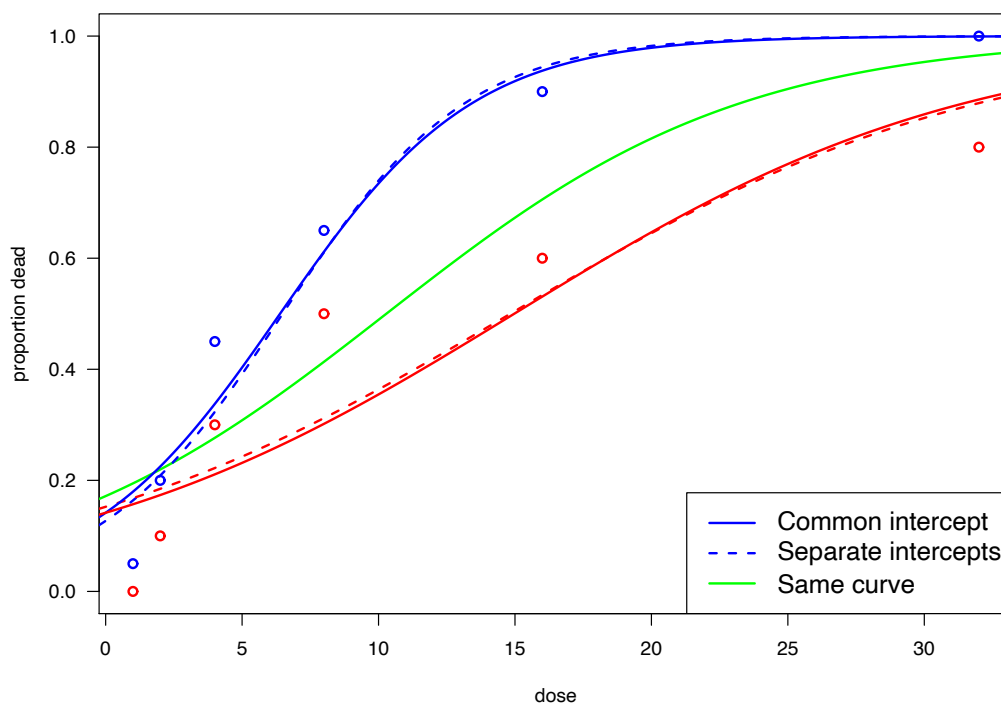
Analysis in R (continued)

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   z-val   P-val
  (Intercept)    -1.80  0.25  -7.17   <0.001
  dose            0.12  0.02   5.63   <0.001
  dose:sexmale    0.16  0.04   3.65   <0.001
```
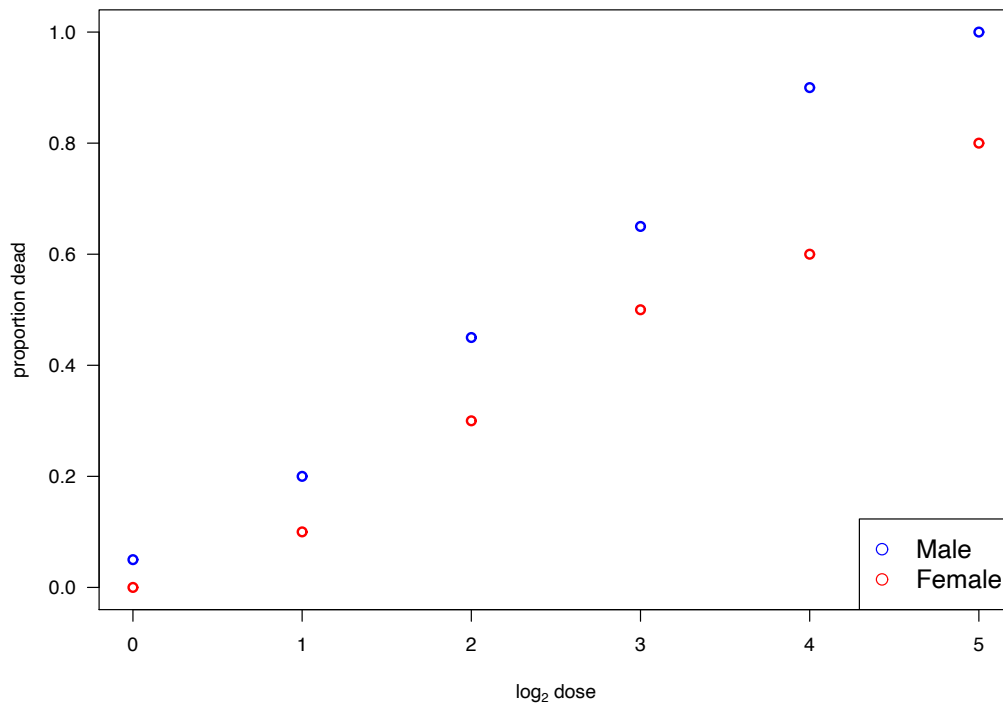
Fitted curves

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   z-val    P-val
  (Intercept)  -2.77  0.37  -7.47  <0.001
  dose          1.01  0.12   8.15  <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    z-val    P-val
  (Intercept)  -2.99  0.55  -5.41   <0.001
  sexmale       0.17  0.78  -0.22    0.822
  dose          0.91  0.17   5.42   <0.001
  sexmale:dose  0.35  0.27   1.31    0.191
```

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   z-val    P-val
  (Intercept)  -2.91  0.39  -7.47   <0.001
  dose          0.88  0.13   6.92   <0.001
  dose:sexmale  0.41  0.12   3.26    0.001
```

Fitted curves