Solutions for 8.6 Exercises

1. **Aspirin and stroke**

(a)
```
> stroke <- data.frame(treat=c("aspirin", "control"),y=c(63,43), n=c(78, 77))
> stroke$rate <- stroke$y/stroke$n
> stroke
    treat  y  n      rate
1 aspirin 63 78 0.8076923
2 control 43 77 0.5584416
> aspirin.1 <- glm(rate ~ factor(treat)
+               , family = binomial
+               , weight = n, data=stroke)
> summary(aspirin.1)
Call:
glm(formula = rate ~ factor(treat), family = binomial, data = stroke,
    weights = n)

Deviance Residuals:
[1]  0  0

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)            1.4351     0.2873   4.995 5.88e-07 ***
factor(treat)control  -1.2002     0.3677  -3.264  0.00110 **

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1.1354e+01  on 1  degrees of freedom
Residual deviance: 1.5321e-14  on 0  degrees of freedom
AIC: 13.132

Number of Fisher Scoring iterations: 3
```

There is a significant difference (at the 0.05 significance level) between the rate of stroke for the aspirin and control groups ($P = 0.0011$).

(b)
```
> exp(summary(aspirin.1)$coef[2,1])
[1] 0.3011204
> exp(summary(aspirin.1)$coef[2,1]+1.96*c(-1,1)*summary(aspirin.1)$coef[2,2])
[1] 0.1464681 0.6190666
```

The estimated odds of a stroke in the aspirin group are lower by a factor of 0.30 compared with the control group (95% CI: [0.146, 0.619]). Or, equivalently, the odds of stroke are higher in the control group by a factor of 3.32, compared with the aspirin group (95% CI: [1.62, 6.83]). This interval does not include 1 and so is consistent with the results of Lab 3, Question 6, where the confidence interval for the difference in the proportions did not include 0. Both indicate a significant difference between the aspirin and control groups.

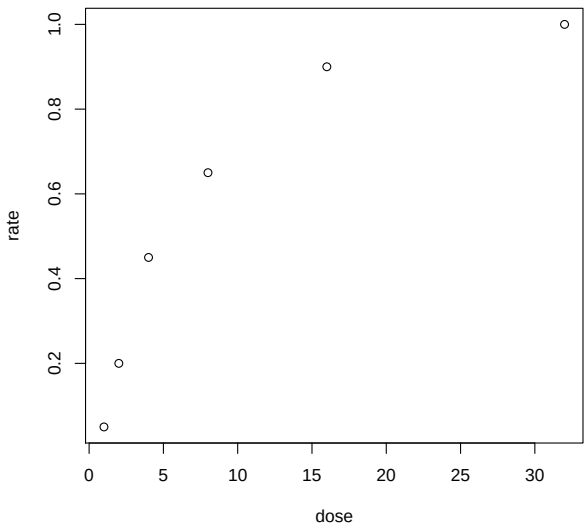
2. **Control of budworm**

(a) Enter the data

```
> moths <- data.frame(dose = c(1, 2, 4, 8, 16, 32),
+ dead = c(1, 4, 9, 13, 18, 20), total = rep(20, 6))
> moths$rate <- moths$dead / 20
```

Firstly, examine a plot of the mortality rate of the moths as a function of dose.

```
> plot(rate ~ dose, data = moths)
```



Fit a logistic regression model and look at the summary:

```
> moths.1 <- glm( rate ~ dose
+               , family = binomial
+               , weight = total
+               , data = moths )
> summary(moths.1)

Call:
glm(formula = rate ~ dose, family = binomial, data = moths, weights = total)

Deviance Residuals:
      1        2        3        4        5        6
-1.5729  -0.0954   1.1798   0.3631  -0.7789   0.1426

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.92771    0.40195  -4.796 1.62e-06 ***
dose         0.29723    0.06254   4.752 2.01e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 71.138  on 5  degrees of freedom
Residual deviance:  4.634  on 4  degrees of freedom
AIC: 22.981

Number of Fisher Scoring iterations: 5
```

The very small P-value for the Wald test indicates that the estimate of the coefficient for dose is significantly different to 0.

(b) `> exp(0.29723)`

```
[1] 1.346125
```

For every unit increase in dose, the odds of a moth dying are estimated to increase by a factor of 1.35, 35%.

(c) The coefficients are:

`> moths.1$coef`

```
(Intercept)        dose
 -1.9277147   0.2972343
```

We solve $0 = -1.92771 + 0.29723\times$ dose

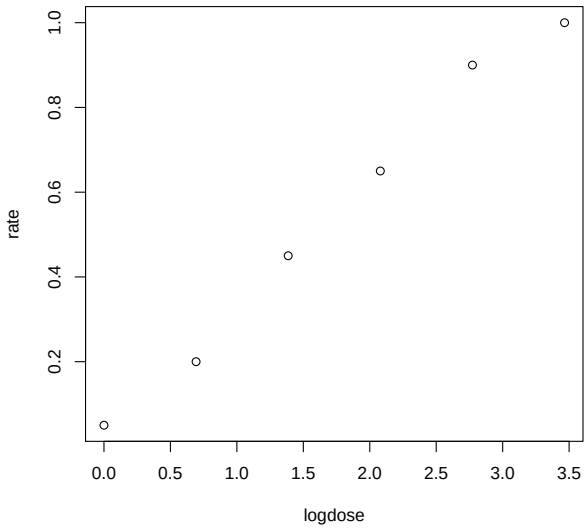
The LD50 is:

`> 1.92771/0.29723`

```
[1] 6.485584
```

(d) `> moths$logdose<-log(moths$dose)`

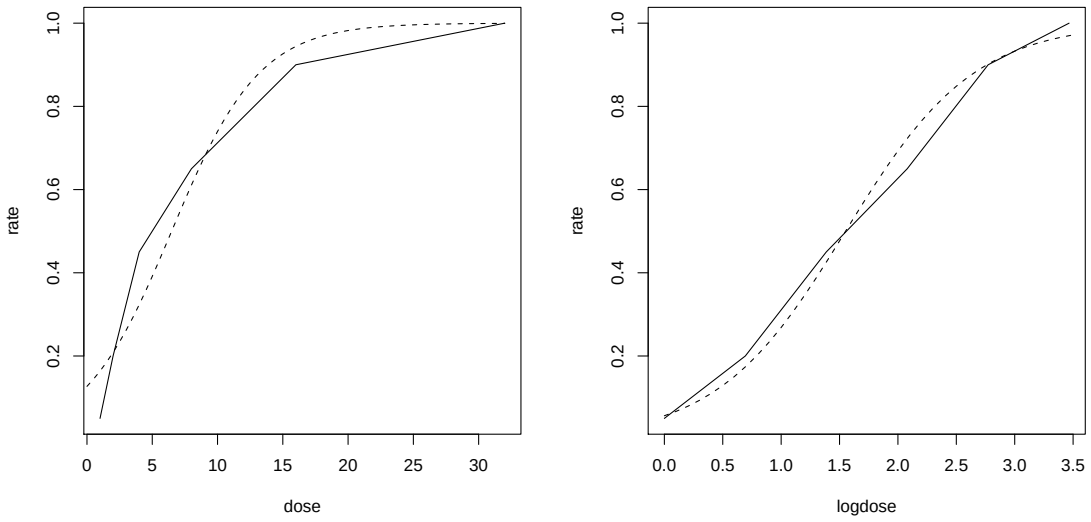
`> plot(rate ~ logdose, data = moths)`



```
> moths.2 <- glm( rate ~ logdose
+               , family = binomial
+               , weight = rep( 20, 6)
+               , data = moths )

> summary(moths.2)

Call:
glm(formula = rate ~ logdose, family = binomial, data = moths,
    weights = rep(20, 6))

Deviance Residuals:
       1         2         3         4         5         6
-0.12505   0.30463   0.22204  -0.71011  -0.02679   1.10375

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.8186     0.5480  -5.143 2.70e-07 ***
logdose       1.8163     0.3059   5.937 2.91e-09 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 71.1376  on 5  degrees of freedom
Residual deviance:  1.8810  on 4  degrees of freedom
AIC: 20.228

Number of Fisher Scoring iterations: 4
```

The P-value is 3 orders of magnitude smaller, and additionally, the residual deviance of 1.88 is smaller than the residual deviance resulting from fitting dose (4.63), so fitting log(dose) provides a better fit. The following plot confirms the better fit:

```
> par(mfrow = c(1, 2))
> plot(rate ~ dose, type = "l", data = moths)
> dse <- seq(0, 32, 0.5)
> lines(dse, predict(moths.1, newdata = data.frame(dose = dse),
+     type = c("response")), lty = 2)
> plot(rate ~ logdose, type = "l", data = moths)
> dse <- seq(0, 3.5, 0.1)
> lines(dse, predict(moths.2, newdata = data.frame(logdose = dse),
+     type = c("response")), lty = 2)
```



3. **Powdery mildew on broccoli**

(a)
```
> broccoli <- data.frame(treat=c(rep("control",5),
+ rep("fungicide1",5), rep("fungicide2",5)),
+ alive=c(3,2,2,1,1,10,8,15,14,8,17,14,13,10,16),
+ emerged=c(17,15,16,11,15,16,11,18,16,12,18,15,16,13,16))
> broccoli$rate <- broccoli$alive/broccoli$emerged
> broccoli
```

```
        treat alive emerged       rate
1     control     3      17 0.17647059
2     control     2      15 0.13333333
3     control     2      16 0.12500000
4     control     1      11 0.09090909
5     control     1      15 0.06666667
6  fungicide1    10      16 0.62500000
7  fungicide1     8      11 0.72727273
8  fungicide1    15      18 0.83333333
9  fungicide1    14      16 0.87500000
10 fungicide1     8      12 0.66666667
11 fungicide2    17      18 0.94444444
12 fungicide2    14      15 0.93333333
13 fungicide2    13      16 0.81250000
14 fungicide2    10      13 0.76923077
15 fungicide2    16      16 1.00000000
```

```
> tapply(broccoli$rate,broccoli$treat,mean)
```

```
   control fungicide1 fungicide2
 0.1184759  0.7454545  0.8919017
```

(b)
```
> broccoli.1 <- glm(rate ~ treat,
+               family = binomial,
+               weight = emerged,
+               data = broccoli )
> summary(broccoli.1)

Call:
glm(formula = rate ~ treat, family = binomial, data = broccoli,
    weights = emerged)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.33513  -0.68944   0.04118   0.68408   1.86087

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.9772     0.3557  -5.559 2.71e-08 ***
treatfungicide1  3.0941     0.4475   6.915 4.69e-12 ***
treatfungicide2  4.1462     0.5155   8.043 8.79e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 127.743  on 14  degrees of freedom
Residual deviance:  12.010  on 12  degrees of freedom
AIC: 53.03

Number of Fisher Scoring iterations: 4
```

The very small P-values show a highly significant difference between the control and each of the two fungicides, consistent with the means from part (a).

(c)
```
> exp(summary(broccoli.1)$coef[2, 1])
[1] 22.0679
> exp(summary(broccoli.1)$coef[2, 1] + 1.96 * c(-1, 1)
+         * summary(broccoli.1)$coef[2, 2])
[1]  9.180438 53.046738
```

4. **Aboriginal deaths in custody**

(a)
```
> deaths <- read.csv("../data/deaths.csv")
> deaths <- data.frame(Year = rep(c(1990,1995),2),
+ Indigenous = rep(c("Yes","No"),each=2),
+ Prisoners = c(2041,2907,12264,14501),
+ Deaths = c(6,17,27,42),
+ Population = c(168317,190438,13141817,13995940))
> deaths$rate <- deaths$Deaths/deaths$Prisoners
> deaths
```

```
  Year Indigenous Prisoners Deaths Population        rate
1 1990        Yes      2041      6     168317 0.002939735
2 1995        Yes      2907     17     190438 0.005847953
3 1990         No     12264     27   13141817 0.002201566
4 1995         No     14501     42   13995940 0.002896352
```

```
> deaths.1<- glm( rate ~ Year
+               , family = binomial
+               , weight = Prisoners
+               , data = deaths )
> summary(deaths.1)
```

```
Call:
glm(formula = rate ~ Year, family = binomial, data = deaths,
    weights = Prisoners)

Deviance Residuals:
      1       2       3       4
 0.5715  2.0662 -0.2450 -1.0476

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -159.61579   86.77194  -1.839   0.0658 .
Year           0.07716    0.04353   1.772   0.0763 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.9920  on 3  degrees of freedom
Residual deviance: 5.7532  on 2  degrees of freedom
AIC: 28.797

Number of Fisher Scoring iterations: 4
```

Although the mortality rate increases from 1990 to 1995 in both groups of prisoners, the increase is not significant at the 0.05 level the P-value from the test.

(b)
```
> deaths.2 <- glm( rate ~ Indigenous
+               , family = binomial
```

```
+               , weight = Prisoners
+               , data = deaths )
> summary(deaths.2)

Call:
glm(formula = rate ~ Indigenous, family = binomial, data = deaths,
    weights = Prisoners)

Deviance Residuals:
      1        2        3        4
-1.2175   0.9140  -0.8434   0.7413

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.9582     0.1205 -49.428   <2e-16 ***
IndigenousYes  0.5916     0.2413   2.452   0.0142 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.9920  on 3  degrees of freedom
Residual deviance: 3.5784  on 2  degrees of freedom
AIC: 26.622

Number of Fisher Scoring iterations: 4
```

From the P-value there was a significantly greater mortality rate for indigenous prisoners at the 0.05 significance level.

(c) The appropriate quantity is an odds ratio. For indigenous vs non-indigenous prisoners:

```
> exp(summary(deaths.2)$coef[2, 1])
[1] 1.806836
> exp(summary(deaths.2)$coef[2, 1] + 1.96 * c(-1, 1) * summary(deaths.2)$coef[2,
+     2])
[1] 1.126017 2.899295
```

Interpretation: for indigenous prisoners, the odds of death in custody are 1.81 times greater than for non-indigenous prisoners (95% CI:[1.13, 2.90]).

(d)
```
> p1 <- 6/2041
> p2 <- 27/12264
> (odds.ratio <- (p1/(1 - p1))/(p2/(1 - p2)))
[1] 1.336282
> (relative.risk <- p1/p2)
[1] 1.335293
```

These are very similar, which is the case when p_1 and p_2 are very small, as they are here.