1. To test the toxicity of a particular substance, six groups of 250 insects each were given different doses of the toxic substance. The dose (x) is given in arbitrary units on a logarithmic scale. One day after exposure, the number of insects that were dead were counted (y).

Use R to perform a logistic regression:

a) Plot the sample proportion that died (y/n) as a function of x. Does the plot suggest that the logistic regression function is appropriate?

From the plot, it appears that a sigmoidal (S-shaped) curve will be appropriate.

b) Find the maximum likelihood estimates for $\beta_0$ and $\beta_1$. State the fitted response function.

```
glm(formula = cbind(died, n - died) ~ Dose, family = binomial(link = "logit"), data = insects)
```

Coefficients:

| Estimate  | Std. Error | z value | Pr(>|z|) |
|-----------|------------|---------|----------|
| (Intercept) | -2.64367   | 0.15610 | -16.93   | <2e-16 |
| Dose       | 0.67399    | 0.03911 | 17.23    | <2e-16 |

2.5 %     97.5 %

| (Intercept) | -2.9554809 | -2.3432165 |
| Dose        | 0.5985828  | 0.7519688  |
Thus, using 95% confidence limits, we have $b_0 = -2.6437 \pm 1.96 \times 0.1561$, and $b_1 = 0.6740 \pm 1.96 \times 0.0391$. The fitted response function is

$$
\ln \left( \frac{\pi}{1 - \pi} \right) = -2.64 + 0.674x
$$

**c)** Plot the sample data together with the fitted response function. Does the fit appear to be a good one?

Yes, the fit appears to be good.

**d)** What is the estimated probability that an insect dies when the dose is $x = 3.5$?

*log-odds:*  
Prediction $\approx -0.2847003$, SE(prediction) $\approx 0.02995912$  
*odds:*  
Odds $\approx 0.7522396$  
*Probability:*  
Probability $\approx 0.4293018$

**e)** What is the estimated median lethal dose (the dose where there is a 50% chance of the insect dying)?

When $\pi = 0.5$, the logistic regression gives $x_{0.5} = -\beta_0/\beta_1$. Thus, the estimated median lethal dose is $2.6437 / 0.6740 \approx 3.922$. 
2. A random sample of 33 families were surveyed to determine their annual family income ($x_1$, in thousands of dollars) as well as the age of their oldest car ($x_2$, in years). One year later, a follow-up survey determined if they bought a car ($y = 1$) or did not purchase a car ($y = 0$) in the past year. Assume that a multiple logistic regression model first order in the two predictor variables is appropriate.

Use R to perform a logistic regression:

- Find the maximum likelihood estimates for $\beta_0$, $\beta_1$, and $\beta_2$. State the fitted response function.

```r
Call: glm(formula = Purchased ~ ., family = binomial(link = "logit"), data = cars)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)  
(Intercept) -4.73931    2.10195  -2.255   0.0242 *  
Income       0.06773    0.02806   2.414   0.0158 *  
CarAge       0.59863    0.39007   1.535   0.1249    

2.5 %  97.5 %
(Intercept) -9.44517202 -1.0486919  
Income       0.01950261  0.1317774  
CarAge      -0.12216880  1.4469327

\[
\ln \left( \frac{\pi}{1 - \pi} \right) = -4.74 + 0.0677 \times income + 0.60 \times car\_age
\]

b) Obtain $\exp(b_1)$ and $\exp(b_2)$ and explain these numbers.

```r
exp(coef(model))

<table>
<thead>
<tr>
<th>(Intercept)</th>
<th>Income</th>
<th>CarAge</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.008744682</td>
<td>1.070079</td>
<td>1.819627</td>
</tr>
</tbody>
</table>
```

The term $\exp(b_1) = 1.07$ is the odds ratio for income. For each $1,000 increase in family income, the odds of buying a new car increase by the factor 1.07. The term $\exp(b_2) = 1.82$ is the odds ratio for car age. For each year increase in the age of the oldest family car, the odds of buying a new car increase by the factor 1.82.

c) What is the estimated probability that a family with an income of $50,000 and an oldest car that is 3 years old will buy a car in the next year?

The predicted probability is 0.609 with a standard error of 0.12.