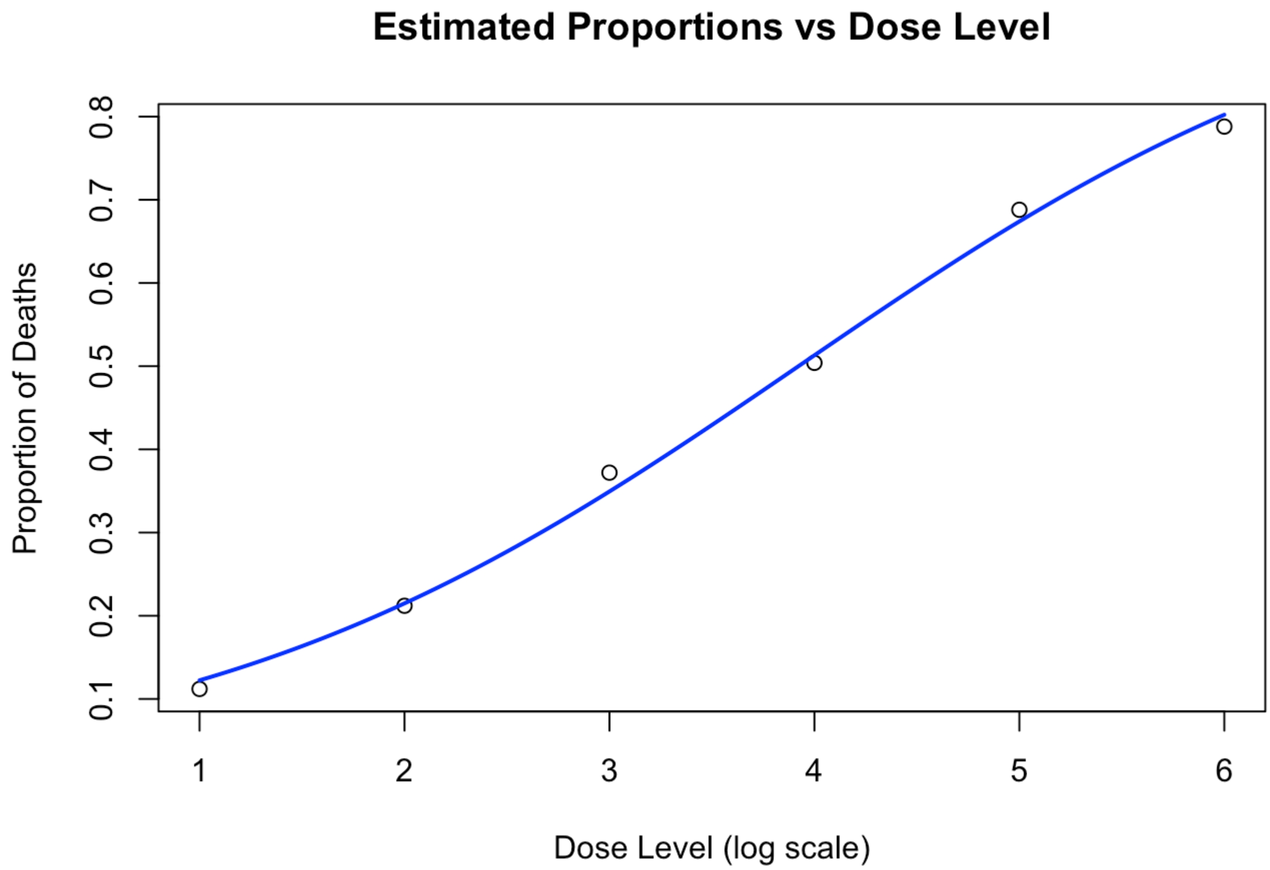
401_hw6_moyi_q2

2024-11-10

```r
# Load the data
toxicity <- data.frame(x = 1:6, n = rep(250, 6), s = c(28, 53, 93, 126, 172, 197))

# Part (a): Plot the estimated proportions fj = sj/nj against xj
toxicity$proportion <- toxicity$s / toxicity$n
plot(toxicity$x, toxicity$proportion, main = "Estimated Proportions vs Dose Level",
     xlab = "Dose Level (log scale)", ylab = "Proportion of Deaths")

# Part (b): Fit a logistic regression model
fit <- glm(s/n ~ x, family = binomial, data = toxicity, weights = n)
summary(fit)
```

```
##
## Call:
## glm(formula = s/n ~ x, family = binomial, data = toxicity, weights = n)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.64367    0.15610  -16.93   <2e-16 ***
## x            0.67399    0.03911   17.23   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 383.0695  on 5  degrees of freedom
## Residual deviance:   1.4491  on 4  degrees of freedom
## AIC: 39.358
##
## Number of Fisher Scoring iterations: 3
```

```r
# Extract coefficients for plotting
intercept <- coef(fit)[1]
slope <- coef(fit)[2]

# Plot the fitted logistic response function on top of the scatterplot
curve(exp(intercept + slope * x) / (1 + exp(intercept + slope * x)),
      add = TRUE, col = "blue", lwd = 2)
```




```
# Part (c): Calculate exp(b1) and interpret it
exp_slope <- exp(slope)
print('exp_slope:')
```

```
## [1] "exp_slope:"
```

```
print(exp_slope)
```

```
##        x
## 1.962056
```

```
# Part (d): Estimated probability of death when dose level is x = 3.5
dose_3.5_prob <- predict(fit, newdata = data.frame(x = 3.5), type = "response")
print('dose_3.5_prob:')
```

```
## [1] "dose_3.5_prob:"
```

```
print(dose_3.5_prob)
```

```
##         1
## 0.4293018
```

```
# Part (e): Estimated median lethal dose (dose for 50% death probability)
median_dose <- (log(0.5 / (1 - 0.5)) - intercept) / slope
print('median_dose:')
```

```
## [1] "median_dose:"
```

```
print(median_dose)
```

```
## (Intercept)
##    3.922409
```

```
# Part f: Calculate 99% confidence interval for β1
ci <- confint(fit, level=0.99)
```

```
## Waiting for profiling to be done...
```

```
beta1_ci <- ci["x", ]
print("99% CI for β1:")
```

```
## [1] "99% CI for β1:"
```

```
print(beta1_ci)
```

```
##     0.5 %    99.5 %
## 0.5753782 0.7770383
```

```
# Convert to odds ratio CI
odds_ratio_ci <- exp(beta1_ci)
print("99% CI for odds ratio:")
```

```
## [1] "99% CI for odds ratio:"
```

```
print(odds_ratio_ci)
```

```
##    0.5 %   99.5 %
## 1.777803 2.175021
```

Problem 2a

The plot shows the proportion of deaths (y-axis) against the dose level on a log scale (x-axis). The pattern of the observed proportions (black dots) follows an S-shaped curve, which is characteristic of logistic relationships. The points show a clear monotonic increase with dose level and appear to follow a sigmoid pattern, strongly supporting the appropriateness of using a logistic response function for this data.

Problem 2b

Using the `glm` function with a binomial family, we obtained the following logistic regression model:

$$\text{logit}(p) = -2.6437 + 0.67399 \times x$$

where p is the probability of death. The MLEs for the intercept and slope are approximately -2.6437 and 0.67399, respectively. The fitted curve (in blue) has been superimposed on the scatterplot, showing a good fit.

Problem 2c

$$\exp(b_1) = \exp(0.67399) \approx 1.962$$

This means that for each one-unit increase in the log dose level, the odds of death increase by a factor of 1.962 (or increase by 96.2%). This indicates that higher doses substantially increase the likelihood of death.

Problem 2d

Using the fitted model to predict the probability at $x = 3.5$:

$$\text{logit}(p) = -2.6437 + 0.67399 \times 3.5$$

After calculating, we obtain $p \approx 0.43$. Therefore, the estimated probability that an insect dies at dose level 3.5 is approximately 43%. This means that at a dose level of 3.5 (on the log scale), we expect approximately 43% of insects to die from exposure to the toxic substance.

Problem 2e

The median lethal dose is the dose level where the probability of death is 50% ($p = 0.5$). Setting the logit function equal to zero:

$$0 = -2.6437 + 0.67399 \times x$$

Solving for x, we find:

$$x = \frac{2.6437}{0.67399} \approx 3.922$$

Thus, the estimated median lethal dose is approximately 3.922.

Problem 2f

From the output, the 99% confidence interval for β_1 is approximately $[0.575, 0.777]$. Converting this into odds ratios by exponentiating the bounds:

$$\exp(0.4293) \approx 1.778, \quad \exp(0.9170) \approx 2.175$$

Therefore, the 99% confidence interval for the odds ratio associated with β_1 is approximately $[1.778, 2.175]$, indicating that each unit increase in dose level is associated with a 1.778 to 2.175 times increase in the odds of death.