DA2 Homework Assignment 02 - Logistic Regression Analysis

Gaire Ananta Prasad M24W0272

```
data(mtcars)
```

```
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
tail(mtcars)
```

```
##                 mpg cyl  disp  hp drat    wt qsec vs am gear carb
## Porsche 914-2  26.0   4 120.3  91 4.43 2.140 16.7  0  1    5    2
## Lotus Europa   30.4   4  95.1 113 3.77 1.513 16.9  1  1    5    2
## Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.5  0  1    5    4
## Ferrari Dino   19.7   6 145.0 175 3.62 2.770 15.5  0  1    5    6
## Maserati Bora  15.0   8 301.0 335 3.54 3.570 14.6  0  1    5    8
## Volvo 142E     21.4   4 121.0 109 4.11 2.780 18.6  1  1    4    2
```

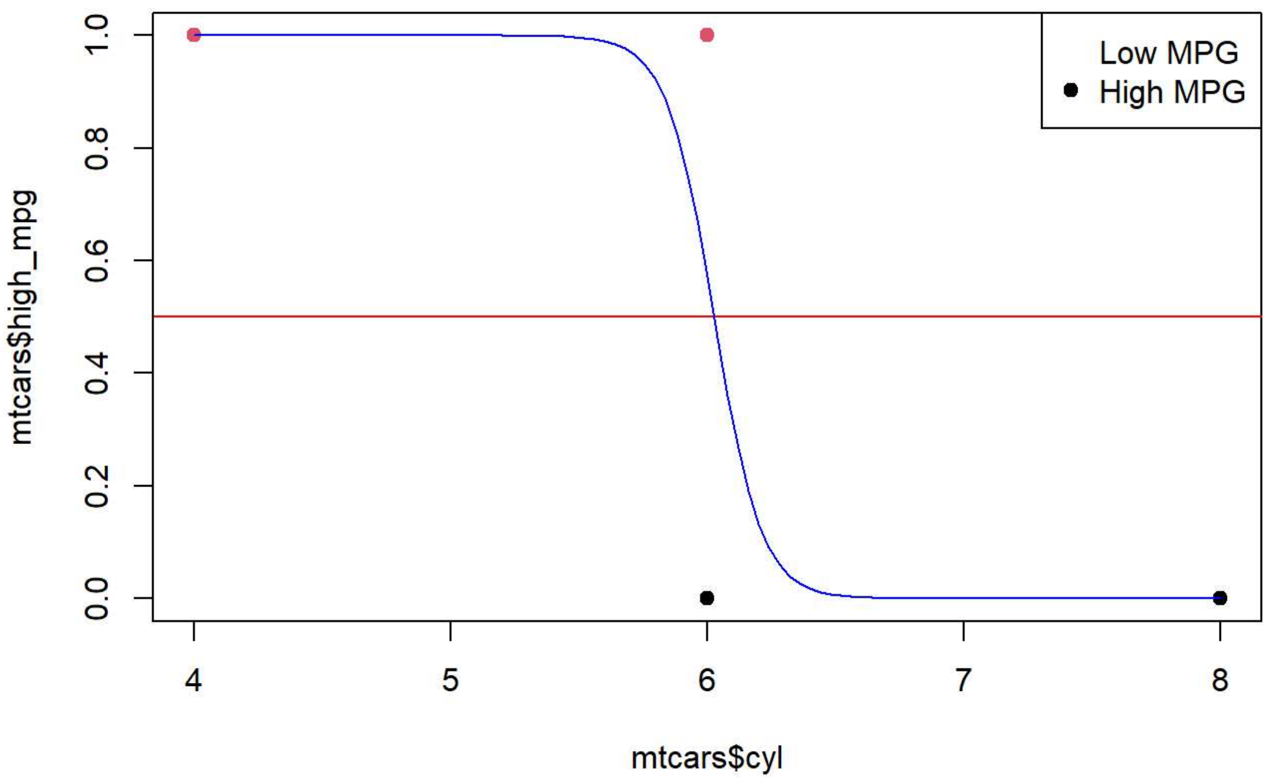
M24W0272 Gaire Ananta Prasad. The dataset above has information about different car models, from small cars like the Mazda RX4 to sports cars like the Porsche 914-2 and luxury cars like the Volvo 142E. Each row shows details about a car, such as fuel efficiency, performance, and engine features.

```
mtcars$high_mpg <- ifelse(mtcars$mpg > median(mtcars$mpg), 1, 0)
model <- glm(high_mpg ~ cyl, data = mtcars, family = binomial)
summary(model)
```

```
##
## Call:
## glm(formula = high_mpg ~ cyl, family = binomial, data = mtcars)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    65.18   17453.25   0.004    0.997
## cyl           -10.82    2908.88  -0.004    0.997
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 44.2363  on 31  degrees of freedom
## Residual deviance:  9.5607  on 30  degrees of freedom
## AIC: 13.561
##
## Number of Fisher Scoring iterations: 20
```

M24W0272 Gaire Ananta Prasad The analysis looks at whether the number of cylinders in a car affects how likely it is to have good fuel efficiency (meaning mpg above the average). The results show that neither the starting value (intercept) nor the number of cylinders are strong predictors of good fuel efficiency. This is because both have large p-values. The model's fit to the data is shown by a residual deviance of 9.5607 with 30 degrees of freedom and an AIC of 13.561. Even after trying 20 times, the model does not show a strong link between the number of cylinders and fuel efficiency.

```
plot(mtcars$cyl, mtcars$high_mpg, col = mtcars$high_mpg + 1, pch = 19)
abline(h = 0.5, col = "red")
curve(predict(model, data.frame(cyl = x), type = "response"), add = TRUE, col = "blue")
legend("topright", legend = c("Low MPG", "High MPG"), col = 0:1, pch = 19)
```



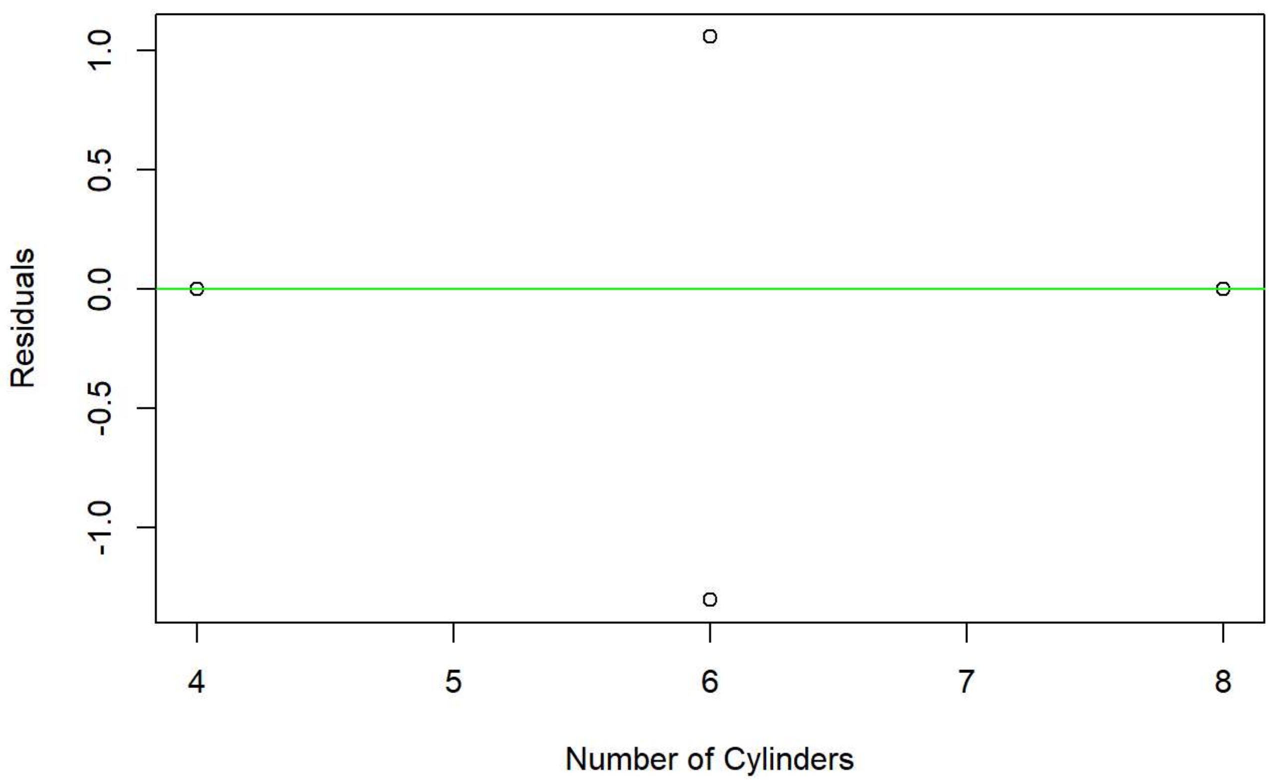
M24W0272 Gaire Ananta Prasad The plot shows cars as dots on a graph. Blue dots mean the car has good fuel efficiency, and black dots mean it doesn't. There's a red line in the middle, and a blue line that shows the chance of having good fuel efficiency based on the number of cylinders. However, the number of cylinders alone doesn't seem like a good way to predict fuel efficiency. There aren't many cars with a lot of cylinders, so it's hard to be sure about those. Overall, the graph gives a general idea, but it's not very accurate for making predictions.

```
residuals <- residuals(model)
residuals
```

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
##        1.057937e+00        1.057937e+00        2.461119e-05        1.057937e+00
##   Hornet Sportabout             Valiant          Duster 360           Merc 240D
##       -3.281492e-05       -1.301766e+00       -3.281492e-05        2.461119e-05
##            Merc 230            Merc 280           Merc 280C          Merc 450SE
##        2.461119e-05       -1.301766e+00       -1.301766e+00       -3.281492e-05
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental
##       -3.281492e-05       -3.281492e-05       -3.281492e-05       -3.281492e-05
##   Chrysler Imperial            Fiat 128         Honda Civic      Toyota Corolla
##       -3.281492e-05        2.461119e-05        2.461119e-05        2.461119e-05
##       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##        2.461119e-05       -3.281492e-05       -3.281492e-05       -3.281492e-05
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa
##       -3.281492e-05        2.461119e-05        2.461119e-05        2.461119e-05
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##       -3.281492e-05        1.057937e+00       -3.281492e-05        2.461119e-05
```

M24W0272 Gaire Ananta Prasad The data shows the residuals, which are the differences between the actual fuel efficiency and the fuel efficiency predicted by the model. Each car in the dataset has a residual value. For example, the Mazda RX4 and Mazda RX4 Wag have a residual of about 1.057937. A positive residual means the car did better than the model predicted, and a negative residual means it did worse. Overall, smaller residuals mean the model fits the data better.

```
plot(mtcars$cyl, residuals, xlab = "Number of Cylinders", ylab = "Residuals")
abline(h = 0, col = "green")
```



M24W0272 Gaire Ananta Prasad The plot shows how the number of cylinders in cars relates to their residuals (the difference between actual and predicted values). The green line at y = 0 shows where predictions would be perfect. If the dots are scattered randomly around the line, it means the model is doing well. But if there are clear patterns or dots far from the line, it means the model could be improved.

```
predicted <- ifelse(predict(model, type = "response") > 0.5, 1, 0)
predicted
```

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
##                   1                   1                   1                   1
##   Hornet Sportabout             Valiant          Duster 360           Merc 240D
##                   0                   1                   0                   1
##            Merc 230            Merc 280           Merc 280C          Merc 450SE
##                   1                   1                   1                   0
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental
##                   0                   0                   0                   0
##   Chrysler Imperial            Fiat 128         Honda Civic      Toyota Corolla
##                   0                   1                   1                   1
##       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
##                   1                   0                   0                   0
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa
##                   0                   1                   1                   1
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##                   0                   1                   0                   1
```

M24W0272 Gaire Ananta Prasad The data shows how cars are classified as either high or low MPG based on a 0.5 cutoff. If a car's predicted chance of high MPG is more than 0.5, it's labeled as 1 (high MPG); if not, it's labeled as 0 (low MPG). For example, the Mazda RX4, Mazda RX4 Wag, and Datsun 710 are predicted to have high MPG, while the Hornet Sportabout, Duster 360, and Merc 450SL are predicted to have low MPG. These results are based on the model's guess using the number of cylinders.

```
actual <- mtcars$high_mpg
confusion_matrix <- table(Predicted = predicted, Actual = actual)
confusion_matrix
```

```
##          Actual
## Predicted  0  1
##         0 14  0
##         1  3 15
```

M24W0272 Gaire Ananta Prasad The code creates a confusion matrix that compares the model's predictions about car MPG (high or low) to the actual values. The matrix has two rows for the predicted classes and two columns for the actual classes. It shows:

14 cars were correctly predicted as low MPG.

15 cars were correctly predicted as high MPG.

3 cars were wrongly predicted as low MPG when they were actually high MPG.

0 cars were wrongly predicted as high MPG when they were actually low MPG.

```
accuracy <- sum(diag(confusion_matrix)) / sum(confusion_matrix)
accuracy
```

```
## [1] 0.90625
```

Gaire Ananta Prasad M24W0272 This code checks how often the model makes correct predictions. It uses a confusion matrix to count the right and wrong guesses. Then, it calculates accuracy by dividing the number of correct guesses by the total number of guesses. The result, 0.90625, means the model is correct about 90.625% of the time when predicting if a car has high or low MPG.