Homework 11

Miles Tweed

5/3/2021

Problem 1

Part 1

$$\begin{aligned}
p(X) &= \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} && (1)\\
1-p(X) &= \frac{1+e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} - \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} && (2)\\
&= \frac{1}{1+e^{\beta_0+\beta_1 X}} && (3)\\
\log\left(\frac{p(X)}{1-p(X)}\right) &= \log\left(\frac{\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{\frac{1}{1+e^{\beta_0+\beta_1 X}}}\right) && (4)\\
&= \log\left(\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \cdot (1+e^{\beta_0+\beta_1 X})\right) && (5)\\
&= \log e^{\beta_0+\beta_1 X} = \beta_0 + \beta_1 X && (6)
\end{aligned}$$

Part 2

a)

$$\begin{aligned}
\frac{p(X)}{1-p(X)} &= 0.37\\
p(X) &= 0.37 - 0.37p(X)\\
p(X) + 0.37p(X) &= 0.37\\
p(X)(1.37) &= 0.37\\
p(X) &= \frac{0.37}{1.37} \approx 0.270
\end{aligned}$$

b)

$$\begin{aligned}
p(X) &= 0.16\\
\frac{p(X)}{1-p(X)} &= \frac{0.16}{1-0.19} = \frac{0.16}{0.84} \approx 0.19
\end{aligned}$$

Problem 2

Part 1

$$\log\left(\frac{\hat{p}(X_1, X_2)}{1-\hat{p}(X_1, X_2)}\right) = -6 + 0.05X_1 + X_2$$
$$\text{where } \hat{p}(X_1, X_2) = \hat{p}(Y = \text{recieve an A} \mid X_1 = \text{Hours Studied and } X_2 = \text{Undergrad GPA})$$

Part 2

$$\begin{aligned}
\log\left(\frac{\hat{p}(x)}{1-\hat{p}(X)}\right) &= -6 + 0.05(40) + 3.5 = -0.5 \\
\frac{\hat{p}(X)}{1-\hat{p}(X)} &= e^{-0.5} = 0.6065 \\
\hat{p}(X) &= 0.6065 - 0.6065\hat{p}(X) \\
\hat{p}(X) &= \frac{0.6065}{1.6065} \approx 0.378
\end{aligned}$$

Part 3

$$\begin{aligned}
\hat{p}(X) &= 0.5 \\
\log\left(\frac{0.5}{0.5}\right) &= 0 = -6 + 0.05X_1 + 3.5 \\
0 &= -2.5 + 0.05X_1 \\
X_1 &= \frac{2.5}{0.05} = 50 \text{ hours}
\end{aligned}$$

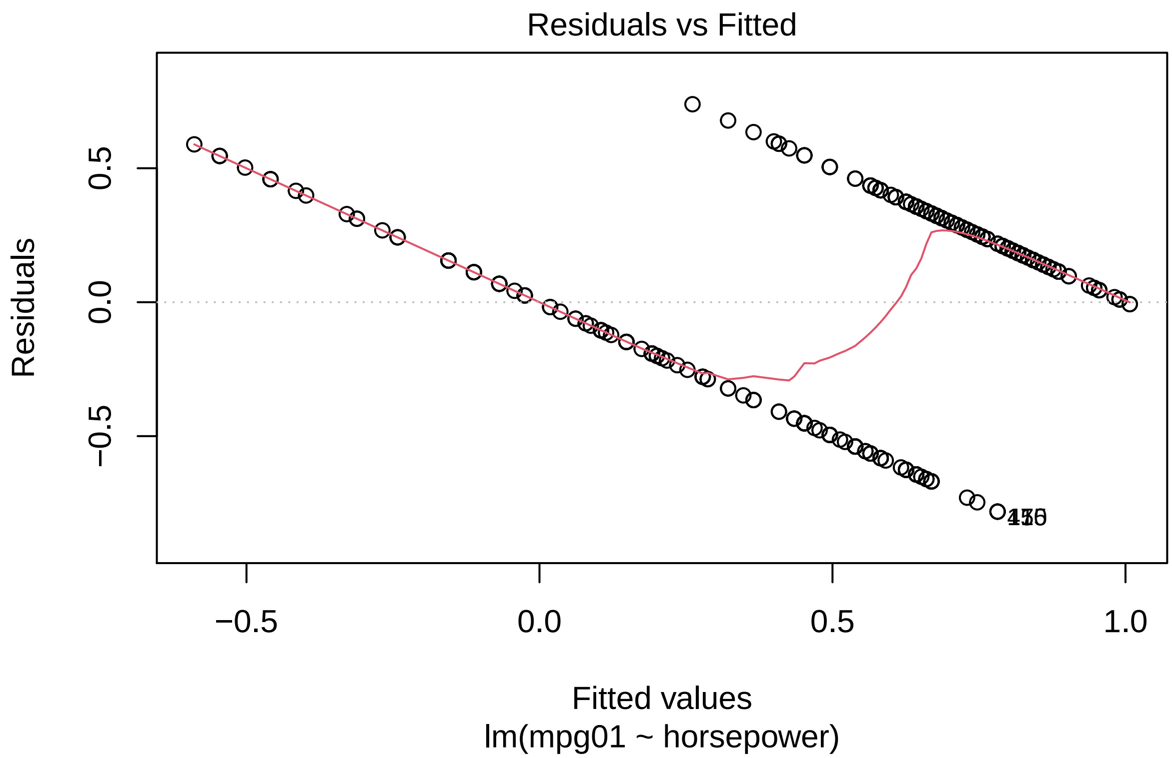
Problem 3

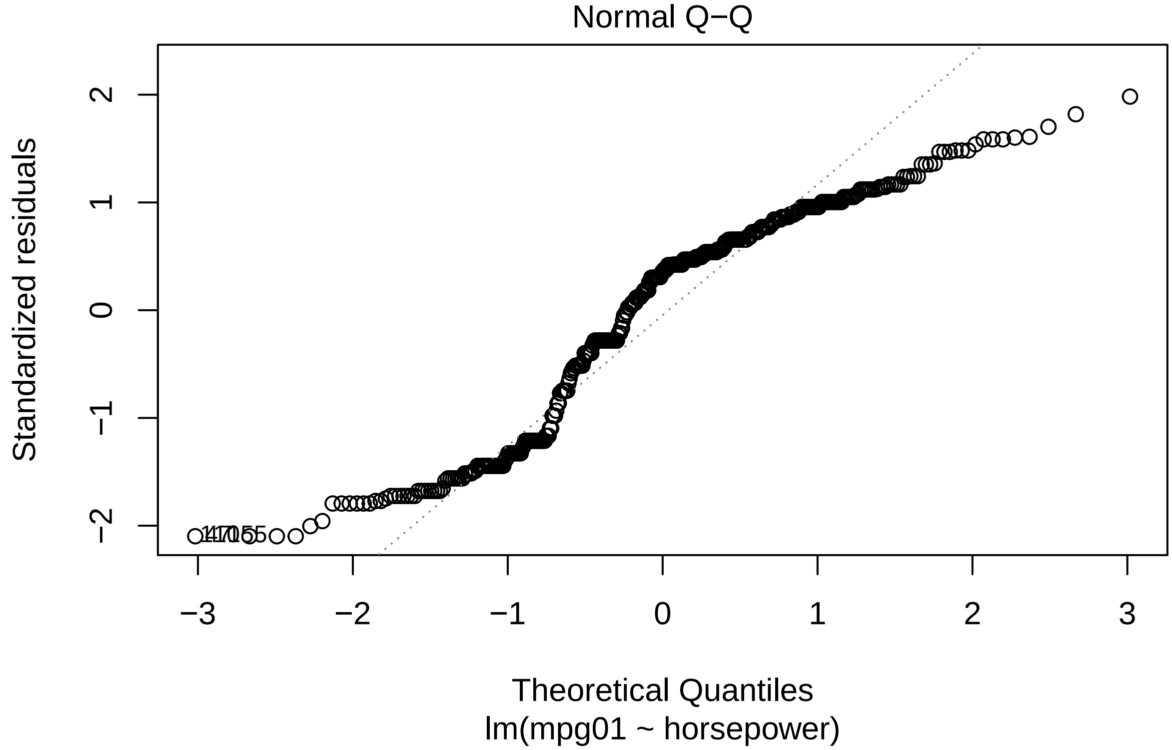
```
library(ISLR)
Auto$mpg01 <- ifelse(Auto$mpg > median(Auto$mpg), 1, 0)
Auto$mpg <- NULL
```

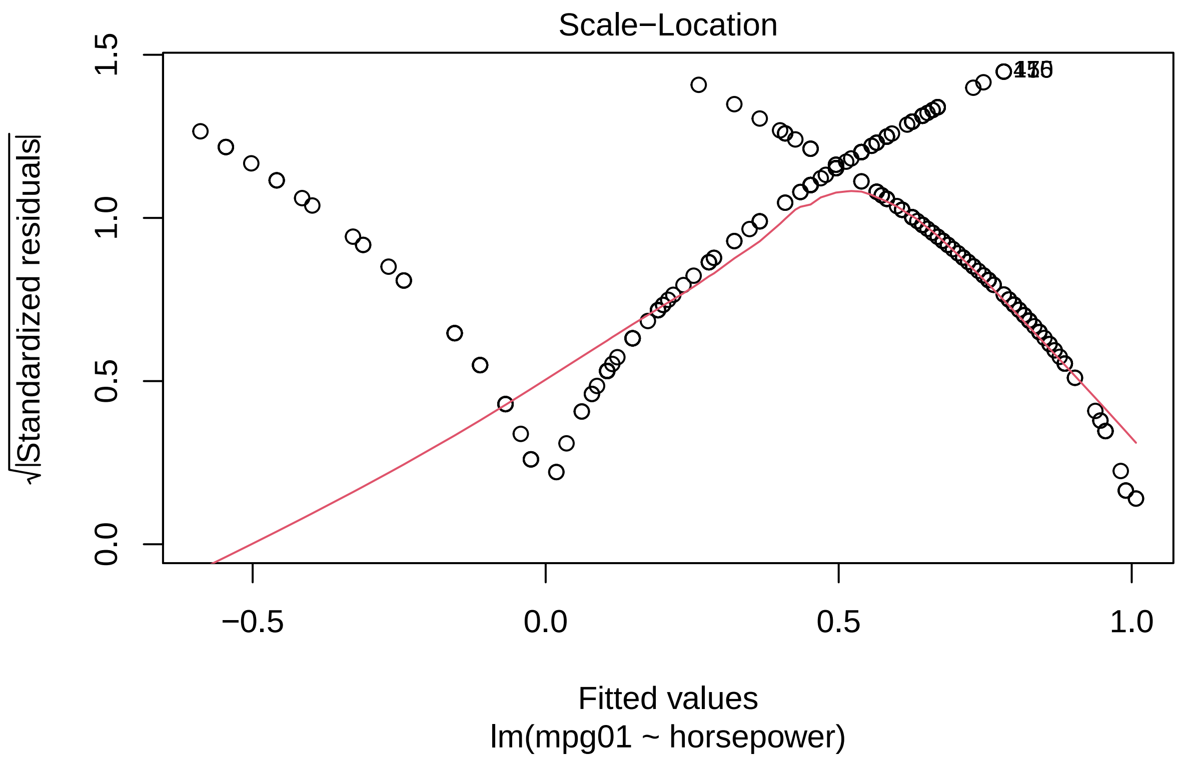
Part1

$$mpg01_i = \beta_0 + \beta_1 horsepower_i + \epsilon_i,\ \epsilon_i \sim \mathbb{N}(0, \sigma^2)$$

```
lm.obj <- lm(mpg01~horsepower, Auto)
plot(lm.obj,which = c(1,2,3))
```


Residuals vs Fitted
Residuals
0.5
0.0
−0.5
−0.5
0.0
0.5
1.0
Fitted values
lm(mpg01 ~ horsepower)


Normal Q−Q
Standardized residuals
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(mpg01 ~ horsepower)



The fitted values are over or under estimates in nearly every case and the error terms are not normally distributed. The linear model attempts to fit continuous values to a response that we know to be binomially distributed which is apparent in the residuals versus fitted plot.

Part 2

a) Let $p_i = p(mpg01_i = 1|horsepower_i)$

$$\begin{cases} mpg_i & \sim_{indep.} Bin(1, p_i), \\ \log\left(\frac{p_i}{1-p_i}\right) & = \beta_0 + \beta_1 horsepower_i \end{cases}$$

b)

```
glm.obj <- glm(mpg01~horsepower, Auto, family = "binomial")
summary(glm.obj)
```

```
##
## Call:
## glm(formula = mpg01 ~ horsepower, family = "binomial", data = Auto)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2025  -0.3445   0.0614   0.5228   2.6467
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  9.29885    1.00692   9.235   <2e-16 ***
## horsepower  -0.09675    0.01080  -8.958   <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: 543.43  on 391  degrees of freedom
## Residual deviance: 276.51  on 390  degrees of freedom
## AIC: 280.51
##
## Number of Fisher Scoring iterations: 6
```

$$\log\left(\frac{\hat{p}(horsepower)}{1-\hat{p}(horsepower)}\right) = 9.299 - 0.0968 horsepower$$
$$\text{where } \hat{p}(X) = \hat{p}(mpg01 = 1 \mid horsepower)$$

c) The relationship is statistically significant with a p-value $< 2 \cdot 10^{-16}$. Per one unit increase in horsepower the odds of a vehicle having mpg greater than the median value decreases by a multiple of $e^{-0.0968} = 0.9077$. Simply, as horsepower increases the odds of a vehicle having mpg greater than the median decreases.

d) The 95% confidence interval for the slope is:

$$(9.299 - 1.96 \cdot 1.01, 9.299 + 1.96 \cdot 1.01) = (7.32, 11.28)$$

If a vehicle has zero horsepower, the odds that the vehicle has mpg greater than the median value are between $e^{7.32} \approx 1510$:1 and $e^{11.29} \approx 79221$:1. In other words, it is almost certain that a vehicle with zero horsepower would have mpg greater than the median value.

e)

$$\begin{aligned}
\log\left(\frac{\hat{p}(horsepower)}{1-\hat{p}(horsepower)}\right) &= 9.299 - 0.0968(80) \\
\frac{\hat{p}(horsepower)}{1-\hat{p}(horsepower)} &= e^{1.555} \\
&= 4.735 \\
\hat{p}(horsepower) &= 4.735 - 4.735 \cdot \hat{p}(horsepower) \\
\hat{p}(horsepower) &= \frac{4.735}{5.735} \approx 0.826
\end{aligned}$$

The odds of a vehicle with 80 horsepower having mpg greater than the median is 4.735 to one which corresponds to a probability of 0.826 or 82.6%.

Part 4

a) Let $p_i = p(mpg01_i = 1 | horsepower_i,\ horsepower_i^2)$

$$\begin{cases} mpg_i & \sim_{indep.} Bin(1, p_i), \\ \log\left(\frac{p_i}{1-p_i}\right) & = \beta_0 + \beta_1 horsepower_i + \beta_2 horsepower_i^2 \end{cases}$$

b)

$$H_0 : \beta_2 = 0,\ H_a : \beta_2 \neq 0$$

c)

```
glm.obj <- glm(mpg01~horsepower+I(horsepower^2), Auto, family = 'binomial')
summary(glm.obj)
```

```
##
## Call:
## glm(formula = mpg01 ~ horsepower + I(horsepower^2), family = "binomial",
##     data = Auto)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.29121  -0.37573  -0.00335   0.49332   2.51330
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)     13.4812326  2.8041695   4.808 1.53e-06 ***
## horsepower      -0.1831477  0.0527093  -3.475 0.000511 ***
## I(horsepower^2)  0.0004350  0.0002447   1.777 0.075498 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 543.43  on 391  degrees of freedom
## Residual deviance: 274.49  on 389  degrees of freedom
## AIC: 280.49
##
## Number of Fisher Scoring iterations: 7
```

$$\log\left(\frac{\hat{p}(horsepower, horsepower^2)}{1-\hat{p}(horsepower, horsepower^2)}\right) = 13.481 - 0.183 horsepower + 0.0004 horsepower^2$$

where $\hat{p}(X_1, X_2) = \hat{p}(mpg01 = 1 \mid horsepower, horsepower^2)$

d) The quadratic relationship is statistically significant at the 90% confidence level but not at the 95% confidence level. P-value = 0.075