chpt3.p8

Claudius Taylor

7/17/2019

(a) Use the lm() function to perform a simple linear regression with "mpg" as the response and "horsepower" as the predictor. Use the summary() function to print the results.

```
data(Auto)
fit <- lm(mpg ~ horsepower, data = Auto)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ horsepower, data = Auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.5710  -3.2592  -0.3435   2.7630  16.9240
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 39.935861   0.717499   55.66   <2e-16 ***
## horsepower  -0.157845   0.006446  -24.49   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.906 on 390 degrees of freedom
## Multiple R-squared:  0.6059, Adjusted R-squared:  0.6049
## F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

How strong is the relationship between the predictor and the response ?

R^2 is equal to 0.6059, almost 61% of the variability in "mpg" can be explained using "horsepower".

Is the relationship between the predictor and the response +ve or -ve ?

Since coeficient of "horsepower" is negative, the relationship is also negative.

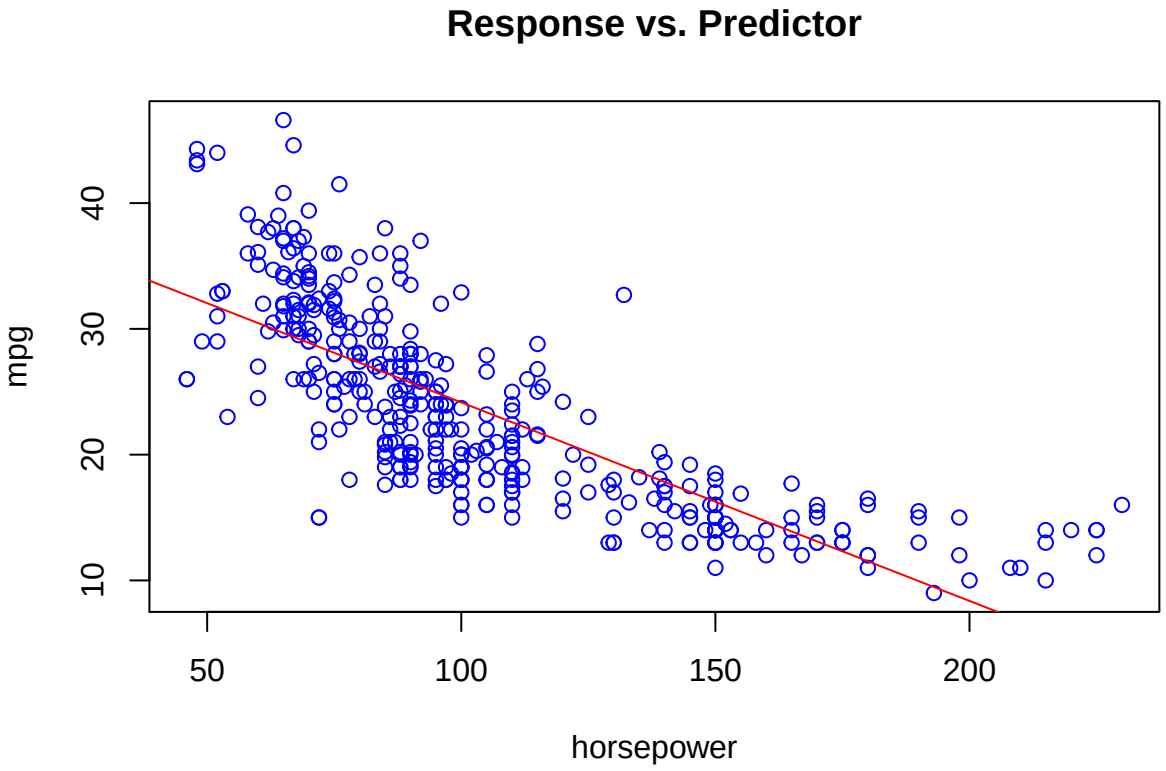
(b) What is the predicted *mpg* associated with a "horsepower" of 98 ? What are the associated 95% confidence and prediction intervals ?

```
predict(fit, data.frame(horsepower = 98), interval = "confidence")
```

```
##        fit      lwr      upr
## 1 24.46708 23.97308 24.96108
```

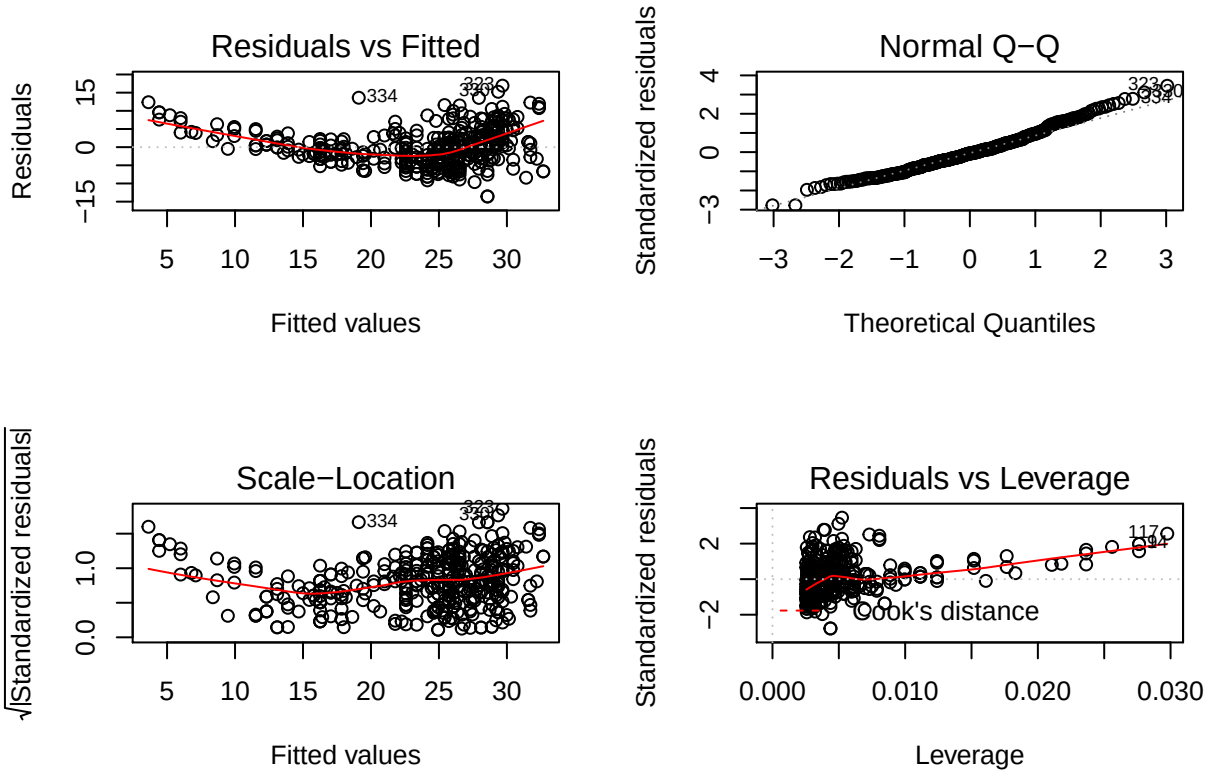
(c) Plot the response and the predictor. Use the abline() function to display the least squares regression line.

```
plot(Auto$horsepower, Auto$mpg, main = "Response vs. Predictor",
     xlab = "horsepower", ylab = "mpg", col = "blue")
abline(fit, col = "red")
```



(d) Use the plot() function to produce diagnostic plots of the least squares regression fit. Comment on any problems you see with the fit.

```
par(mfrow = c(2, 2))
plot(fit)
```



The plot of residuals versus fitted values indicates the presence of nonlinearity in the data. The plot of standardized residuals versus leverage indicates the presence of a few outliers (higher than 2 or lower than -2) and a few high leverage points.