Bruce Campbell ST-617 HW 1

Tue Jun 28 22:40:51 2016

Chapter 3

Problem 8

a)

```
rm(list = ls())
library(ISLR)
library(pander)
DF <- Auto
pander(names(DF))
```

mpg, *cylinders*, *displacement*, *horsepower*, *weight*, *acceleration*, *year*, *origin* and *name*

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

```
##
## Call:
## lm(formula = mpg ~ horsepower, data = DF)
##
## 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
```

```
Yhat <- function(beta0, beta1, predictor) {
    result <- beta0 + beta1 * predictor
    return(unname(result))
}

mpg_h98 <- Yhat(lm.fit$coefficients["(Intercept)"], lm.fit$coefficients["horsepower"],
    98)
```

We note that there is a negative repationship between horespower and mpg. When horsepower goes up mpg goes down. We note that the coefficients β_0 and β_1 are large compared to their standard errors and that the p-value of the t-statistic is very small

The confidence intervals for the regression coefficients are not too 38.5252118, -0.170517, 41.3465103, -0.1451725

The predicted value of mpg for a horespower of 98 is 24.4670772 as calculated by our function above. Using the predict function we get the prediction intervals

```
predict(lm.fit, data.frame(horsepower = c(98)), interval = "prediction")
```

```
##        fit     lwr      upr
## 1 24.46708 14.8094 34.12476
```

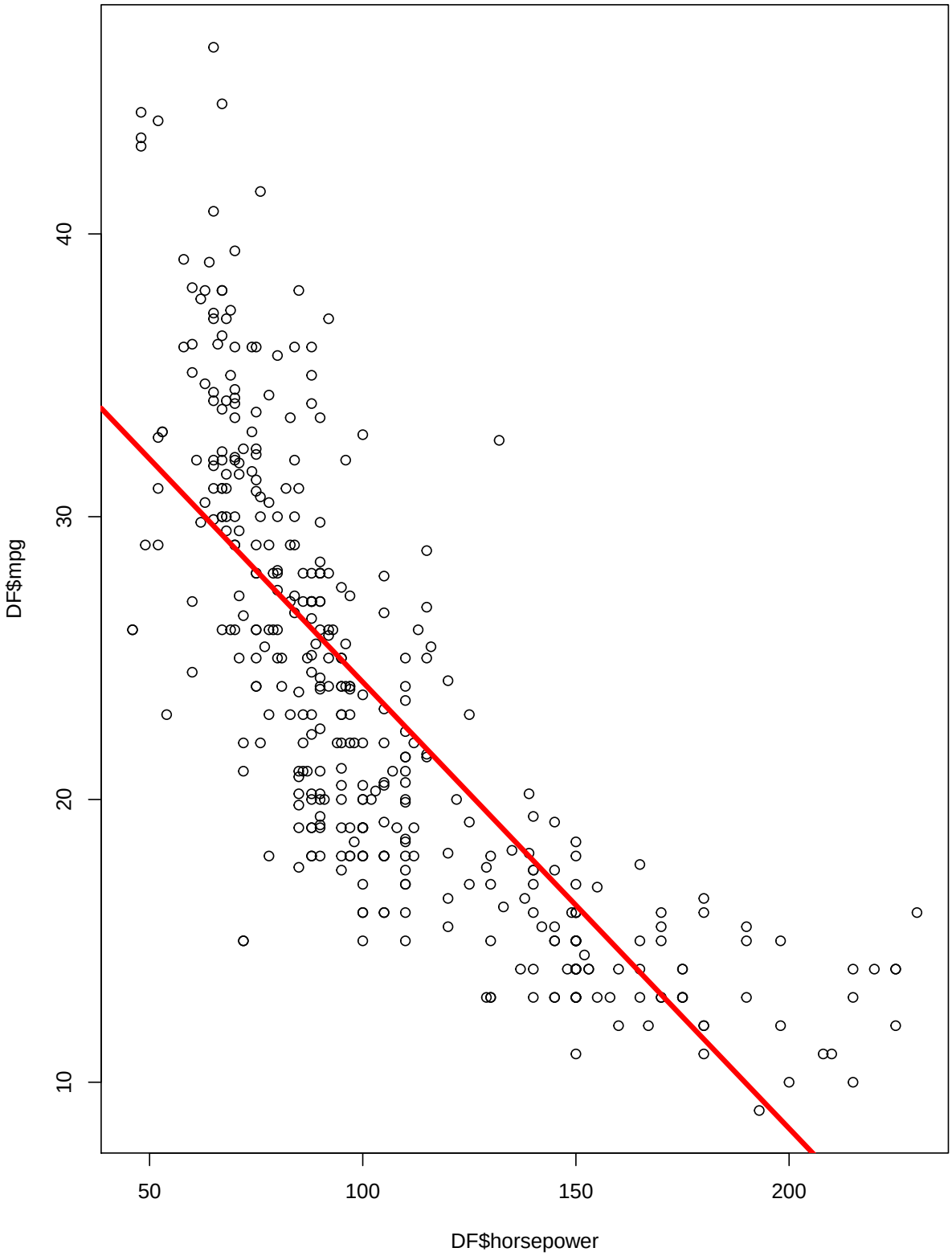
and confidance intervals for this point

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

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

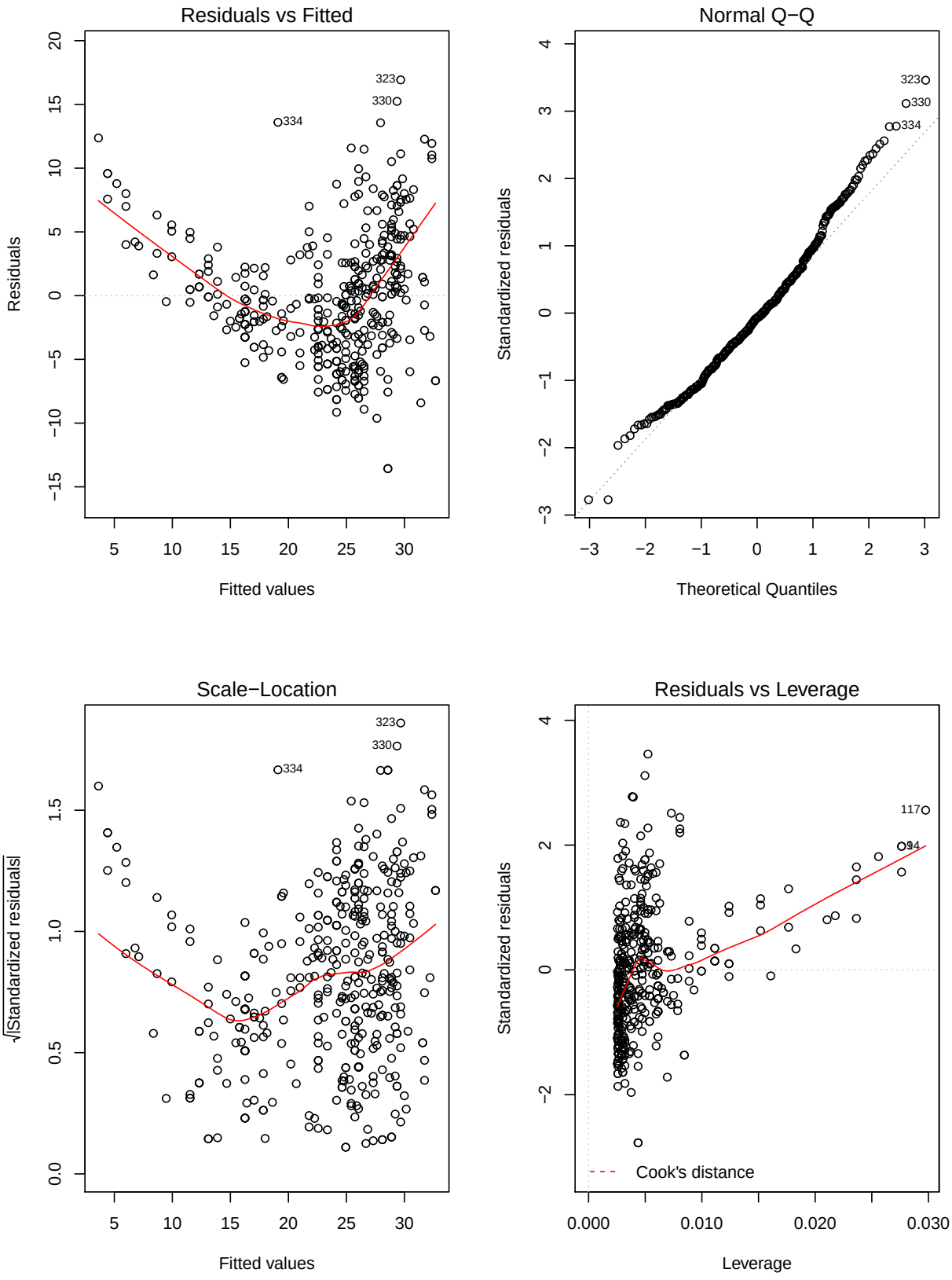
b)

```
plot(DF$horsepower, DF$mpg)
abline(lm.fit, col = "red", lwd = 4)
```


DF$mpg
10
20
30
40
50
100
150
200
DF$horsepower

c) Diagnostic plots

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


Residuals vs Fitted
Residuals
Fitted values
323
330
334
Normal Q−Q
Standardized residuals
Theoretical Quantiles
323
330
334
Scale−Location
√|Standardized residuals|
Fitted values
323
330
334
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
117

From the diagnostic plots we see that there are some high leverage points and that residuals are higher at the boundaries of the range of the perdictor. We also note some divergence from normality for the standardised residuals.