

Regression Midterm 1 Take Home

Chris Kalra

2/21/2019

Your solutions to this midterm must be entirely your own work. Your submission must include a cover sheet with the following statement:

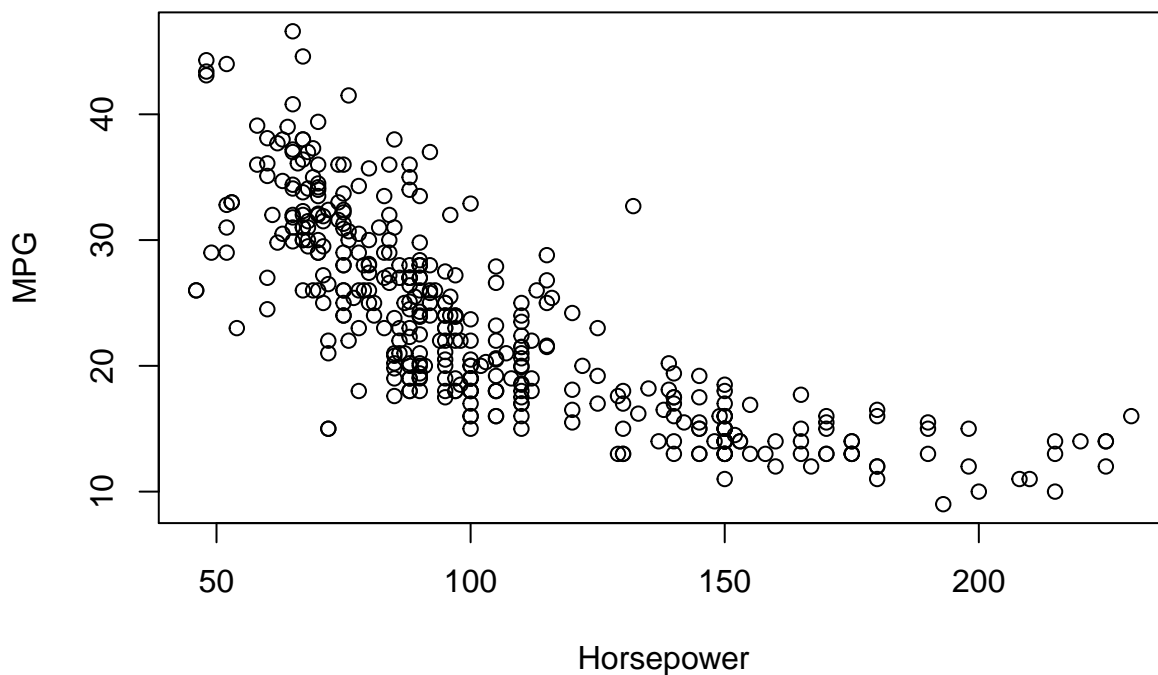
The work contained in this document is entirely my own; I have not consulted with anyone other than the instructor.

Signed,

Chris Kalra

1

```
#a) install.packages("ISLR")
library(ISLR)
plot(Auto$horsepower, Auto$mpg, xlab="Horsepower", ylab="MPG")
```



```
#b)
reg1b=lm(mpg ~ horsepower + I(horsepower^2), data=Auto)
summary(reg1b)
```

```
##
## Call:
## lm(formula = mpg ~ horsepower + I(horsepower^2), data = Auto)
##
## Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -14.7135 -2.5943 -0.0859   2.2868 15.8961
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   56.9000997   1.8004268   31.60  <2e-16 ***
## horsepower    -0.4661896   0.0311246  -14.98  <2e-16 ***
## I(horsepower^2) 0.0012305   0.0001221   10.08  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.374 on 389 degrees of freedom
## Multiple R-squared:  0.6876, Adjusted R-squared:  0.686
## F-statistic:  428 on 2 and 389 DF, p-value: < 2.2e-16
```

```
#c)
x_new = data.frame(horsepower=150)
predict(regr1b, newdata=x_new, interval = "prediction")
```

```
##      fit      lwr      upr
## 1 14.65872 6.027273 23.29016
```

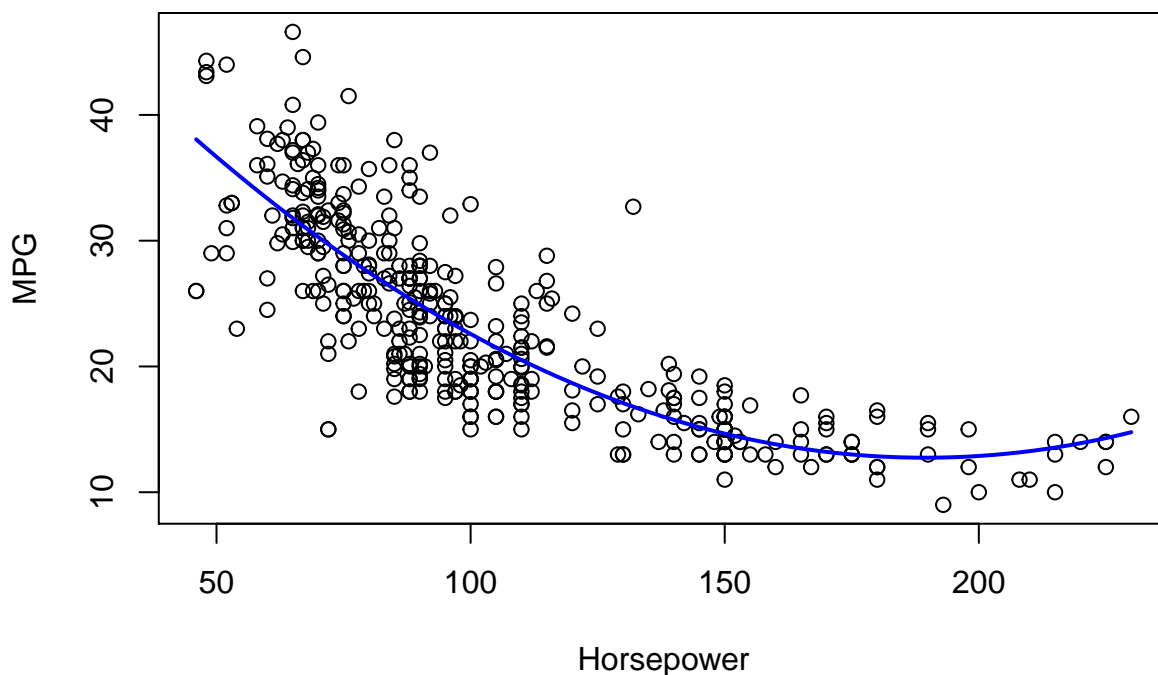
```
# Note that the default prediction interval is a 95% prediction interval
```

```
#d)
range(Auto$horsepower)
```

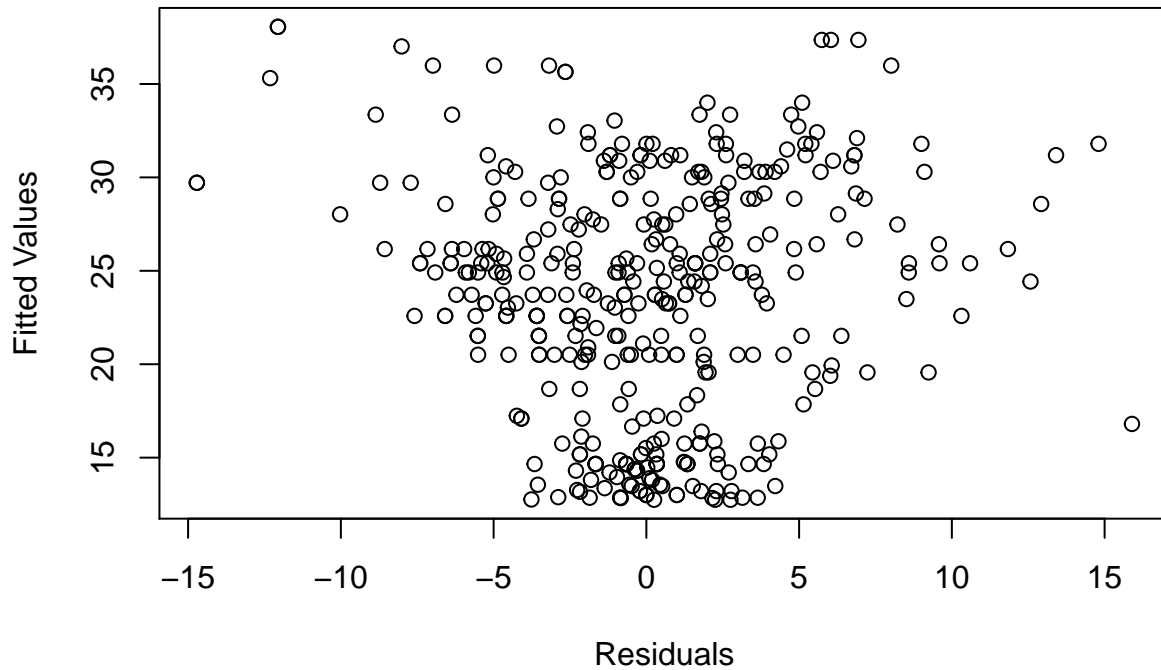
```
## [1] 46 230
```

```
x_grd = seq(46, 230, by=1)
x_new = data.frame(horsepower = x_grd)
preds = predict(regr1b, newdata = x_new)
```

```
plot(mpg~horsepower,data=Auto,ylab='MPG',xlab='Horsepower') ; lines(x_grd,preds, col='blue',lwd=2)
```

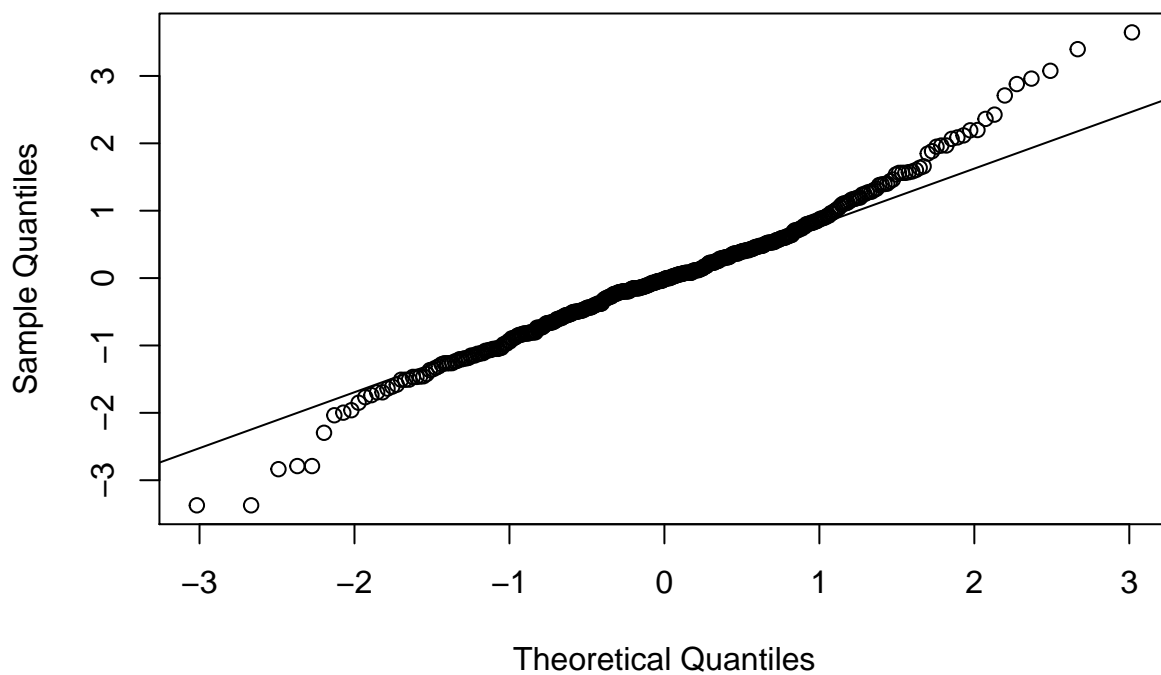


```
#e)
res = resid(regr1b)
fit = fitted.values(regr1b)
plot (res, fit, xlab="Residuals", ylab="Fitted Values")
```



```
stdres = rstandard(regr1b)
qqnorm(stdres) ; qqline(stdres)
```

Normal Q-Q Plot



Unfortunately, the standardized residuals certainly have a “V” pattern to them, which indicates violation of constant variance

Likewise, the QQ plot shows some deviation fairly early on; the upper tail begins to deviate from the line noticeably after 1 SD, while the lower tail begins to deviate from the line at approximately -1.5 SD. Therefore, the assumption of normality may be violated as well