Assignment 2 Economics 573

2022-09-15

Conceptual Exercises

1. The null hypothesis for radio is that in the presence of TV and newspaper ads, radio ads have no effect on sales. The null hypothesis for TV is that in the presence of radio and newspaper ads, TV ads have no effect on sales. The null hypothesis for newpaper is that in the presense of radio and TV ads, newspaper ads have no effect on sales. The low p-values of TV and Radio suggest that their null hypotheses are false and the high p-value of newspaper suggests that the null hypothesis is true.

3a. ii. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates is correct.

3b. 137.1

3c. This is false. To determine if the interaction term is statistically significant or not, we must look at the p-value of the regression coefficient.

4a: We would expect the cubic regression to have a lower training RSS than the linear regression because it could potentially allow for a tighter fit.

4b: We would expect the cubic regression to have a higher test RSS as the overfit would have more error than the linear regression.

4c: We would expect the cubic regression to have a lower training RSS than the linear regression because of the cubic regression's higher flexibility. Since it has higher flexibility, it will follow the point more closely and reduce training error when compared to a linear regression.

4d: Since we don't know how far from linear the relationship is between X and Y, there is not enough information to determine which regression would give a lower test RSS. If closer to linear than cubic, the linear regression would give a lower test RR, but if the inverse was true, it would potentially have a higher test RSS.

Applied Exercises

Question 8a

```
library(ISLR2)
data("Auto")
simple.fit<-lm(mpg~horsepower, data=Auto)
summary(simple.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
```

8ai. Yes, there is a relationship between the predictor and the response since the p-value<2.2e-16.

8aii. The R-squared value indicates that about 61% of the variation in mpg (response variable) is due to horsepower (predictor variable).

8aiii. The relationship between the predictor (horsepower) and the response (mpg) is negative.

8aiv. The predicted mpg associated with a horsepower of 98 is 24.47.

```
q8PredMPG<-predict(simple.fit,data.frame(horsepower=c(98)), interval="prediction")
q8ConfMPG<-predict(simple.fit,data.frame(horsepower=c(98)), interval="confidence")
print(q8PredMPG)
```

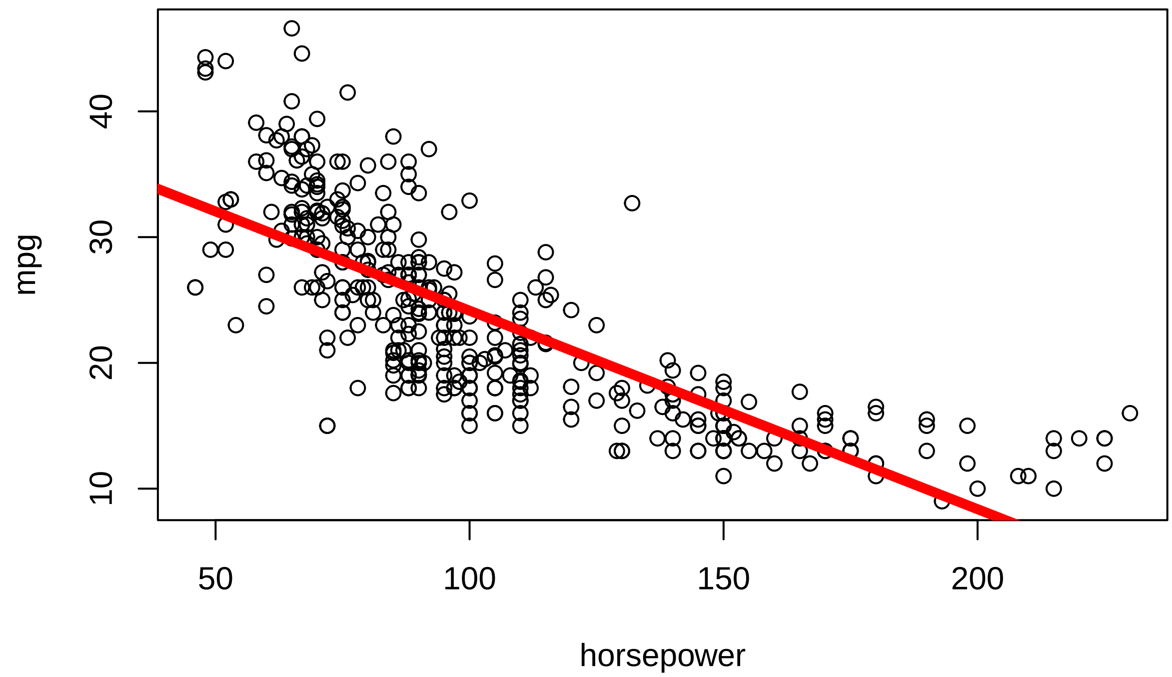
```
##        fit     lwr      upr
## 1 24.46708 14.8094 34.12476
```

```
print(q8ConfMPG)
```

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

Question 8b

```
attach(Auto)
plot(horsepower, mpg)
abline(simple.fit,lwd=5, col="red")
```

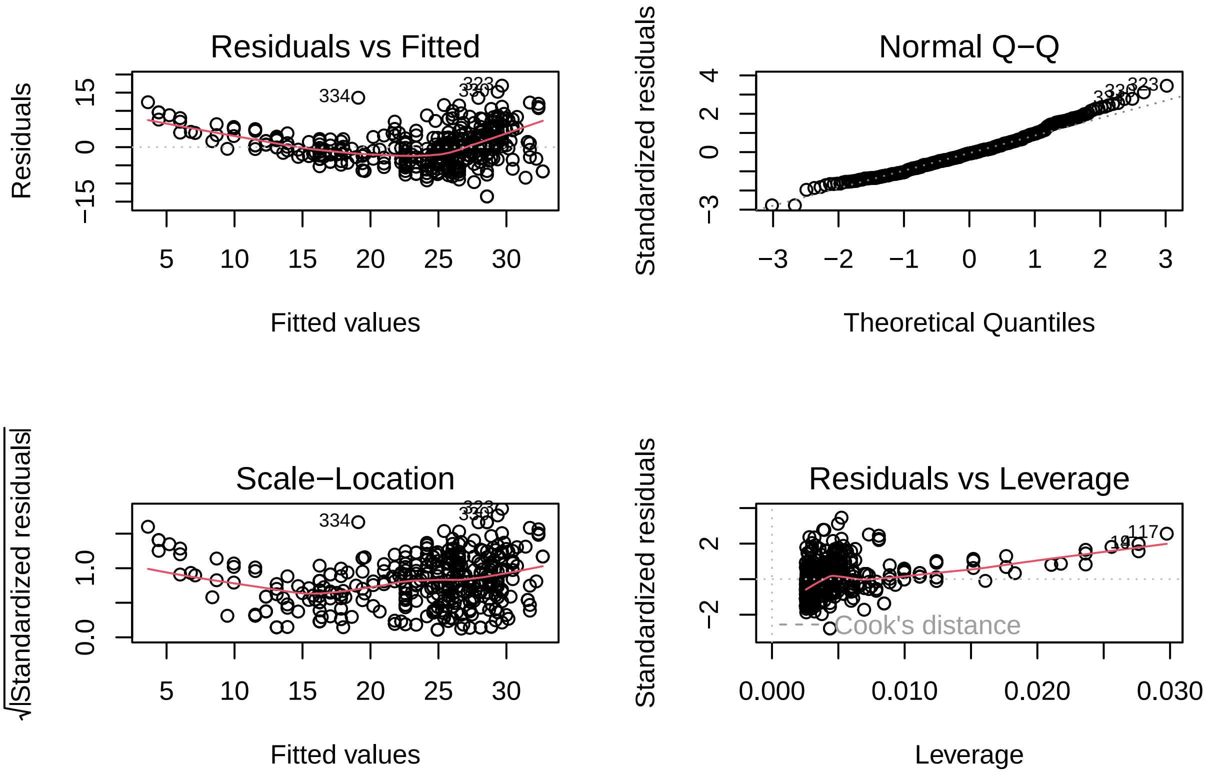


Question 8c

```
which.max(hatvalues(simple.fit))
```

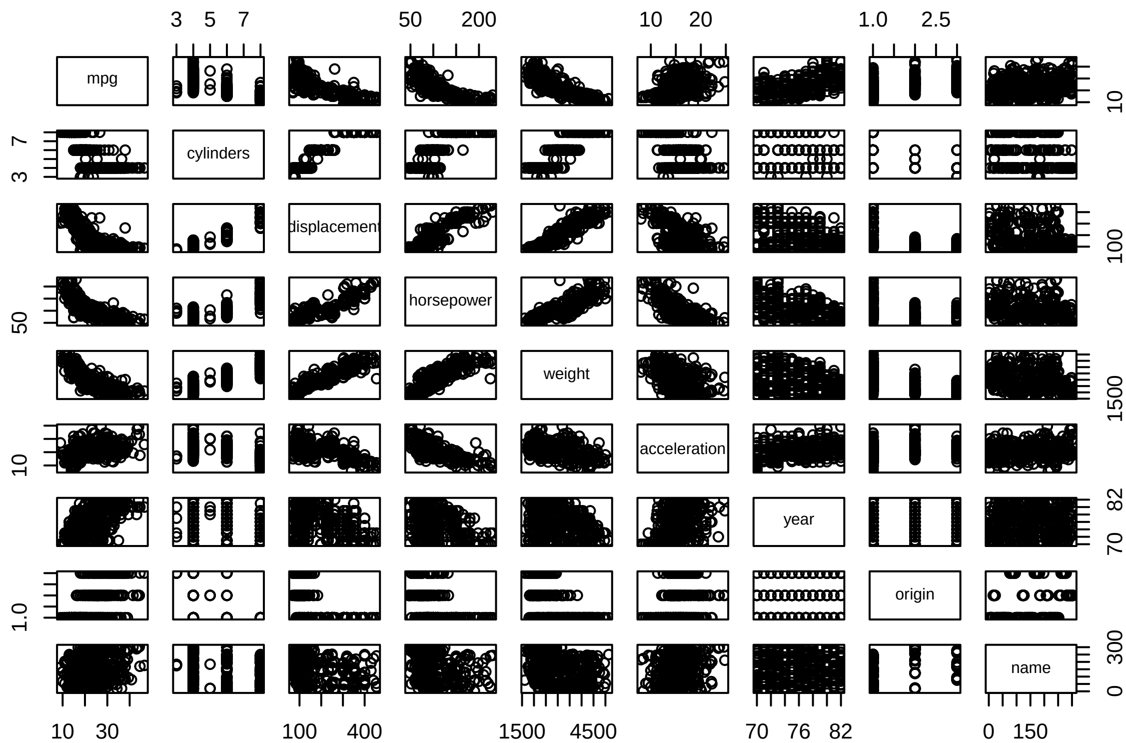
```
## 117
## 116
```

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



Question 9a

```
pairs(Auto)
```



Question 9b

```
Auto$name<-NULL
cor(Auto,method= c("pearson"))
```

```
##                     mpg  cylinders displacement horsepower
## mpg           1.0000000 -0.7776175   -0.8051269 -0.7784268
## cylinders    -0.7776175  1.0000000    0.9508233  0.8429834
## displacement -0.8051269  0.9508233    1.0000000  0.8972570
## horsepower   -0.7784268  0.8429834    0.8972570  1.0000000
## weight       -0.8322442  0.8975273    0.9329944  0.8645377
## acceleration  0.4233285 -0.5046834   -0.5438005 -0.6891955
## year          0.5805410 -0.3456474   -0.3698552 -0.4163615
## origin        0.5652088 -0.5689316   -0.6145351 -0.4551715
##                  weight acceleration       year     origin
## mpg          -0.8322442    0.4233285  0.5805410  0.5652088
## cylinders     0.8975273   -0.5046834 -0.3456474 -0.5689316
## displacement  0.9329944   -0.5438005 -0.3698552 -0.6145351
## horsepower    0.8645377   -0.6891955 -0.4163615 -0.4551715
## weight        1.0000000   -0.4168392 -0.3091199 -0.5850054
## acceleration -0.4168392    1.0000000  0.2903161  0.2127458
## year         -0.3091199    0.2903161  1.0000000  0.1815277
## origin       -0.5850054    0.2127458  0.1815277  1.0000000
```

Question 9c

```r
q9sim.fit<-lm(mpg~.,data=Auto)
summary(q9sim.fit)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5903 -2.1565 -0.1169  1.8690 13.0604
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -17.218435   4.644294  -3.707  0.00024 ***
## cylinders     -0.493376   0.323282  -1.526  0.12780
## displacement   0.019896   0.007515   2.647  0.00844 **
## horsepower    -0.016951   0.013787  -1.230  0.21963
## weight        -0.006474   0.000652  -9.929  < 2e-16 ***
## acceleration   0.080576   0.098845   0.815  0.41548
## year           0.750773   0.050973  14.729  < 2e-16 ***
## origin         1.426141   0.278136   5.127 4.67e-07 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.328 on 384 degrees of freedom
## Multiple R-squared:  0.8215, Adjusted R-squared:  0.8182
## F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

9ci. Yes, there is a relationship between the predictors and the response because the p-value<2.2e-16.

9cii. Displacement, weight, year, and origin appear to have a statistically significant relationship to the response.

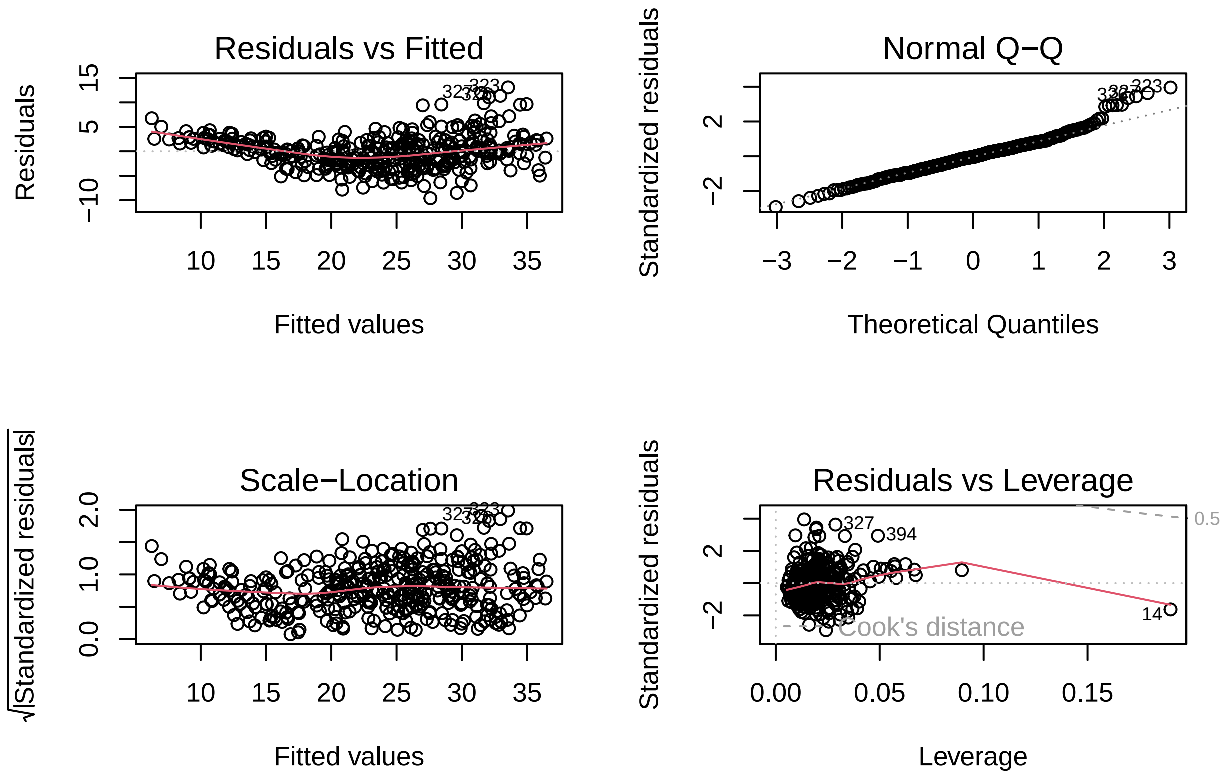
9ciii. The coefficient for the year variable suggests when every other variable held constant, the mpg value increases about 0.75 with each year that passes, meaning as cars get newer, the mpg goes up.

Question 9d

```r
which.max(hatvalues(q9sim.fit))
```

```
## 14
## 14
```

```r
par(mfrow = c(2,2))
plot(q9sim.fit)
```



9d: The Residuals Vs Fitted graph shows that there is a non-linear relationship between the response and predictors. The next graph shows that the residuals are normally distributed and slightly right skewed. The third graph shows that the constant variance of error assumption is not true for this model and the last graph shows that there are no leverage points, but there is an observation that stands out as a potential leverage point.

Question 9e

```
q9sim.fit<-lm(mpg ~.-name+displacement:weight, data = Auto)
```

```
## Warning in terms.formula(formula, data = data): 'varlist'
## has changed (from nvar=8) to new 9 after EncodeVars() --
## should no longer happen!
```

```
summary(q9sim.fit)
```

```
##
## Call:
## lm(formula = mpg ~ . - name + displacement:weight, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.9027 -1.8092 -0.0946  1.5549 12.1687
```

```
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -5.389e+00  4.301e+00  -1.253   0.2109
## cylinders             1.175e-01  2.943e-01   0.399   0.6899
## displacement         -6.837e-02  1.104e-02  -6.193 1.52e-09
## horsepower           -3.280e-02  1.238e-02  -2.649   0.0084
## weight               -1.064e-02  7.136e-04 -14.915  < 2e-16
## acceleration          6.724e-02  8.805e-02   0.764   0.4455
## year                  7.852e-01  4.553e-02  17.246  < 2e-16
## origin                5.610e-01  2.622e-01   2.139   0.0331
## displacement:weight  2.269e-05  2.257e-06  10.054  < 2e-16
##
## (Intercept)
## cylinders
## displacement         ***
## horsepower           **
## weight               ***
## acceleration
## year                 ***
## origin               *
## displacement:weight ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.964 on 383 degrees of freedom
## Multiple R-squared:  0.8588, Adjusted R-squared:  0.8558
## F-statistic: 291.1 on 8 and 383 DF,  p-value: < 2.2e-16
```

9e: Yes, some interactions do appear to be statistically significant.

Question 9f

```
q9sim.fit = lm(mpg ~.-name+I((displacement)^2)+log(displacement)+displacement:weight, data = Auto)
```

```
## Warning in terms.formula(formula, data = data): 'varlist'
## has changed (from nvar=10) to new 11 after EncodeVars() --
## should no longer happen!
```

```
summary(q9sim.fit)
```

```
##
## Call:
## lm(formula = mpg ~ . - name + I((displacement)^2) + log(displacement) +
##     displacement:weight, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7453 -1.8071  0.0077  1.5523 12.2398
##
```

```
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -4.372e+01  2.127e+01  -2.056 0.040508
## cylinders             6.809e-01  3.756e-01   1.813 0.070618
## displacement         -1.965e-01  6.336e-02  -3.101 0.002073
## horsepower           -4.658e-02  1.390e-02  -3.351 0.000886
## weight               -9.389e-03  1.415e-03  -6.633 1.13e-10
## acceleration          4.618e-02  8.993e-02   0.514 0.607885
## year                  7.673e-01  4.596e-02  16.696  < 2e-16
## origin                5.165e-01  2.713e-01   1.904 0.057702
## I((displacement)^2)  1.737e-04  7.263e-05   2.391 0.017291
## log(displacement)     1.046e+01  5.796e+00   1.805 0.071801
## displacement:weight  1.889e-05  4.645e-06   4.067 5.78e-05
##
## (Intercept)          *
## cylinders            .
## displacement         **
## horsepower           ***
## weight               ***
## acceleration
## year                 ***
## origin               .
## I((displacement)^2) *
## log(displacement)    .
## displacement:weight ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.949 on 381 degrees of freedom
## Multiple R-squared:  0.8609, Adjusted R-squared:  0.8572
## F-statistic: 235.7 on 10 and 381 DF,  p-value: < 2.2e-16
```

Question 13a

```
set.seed(1)
x<-rnorm(100)
```

Question 13b

```
eps<-rnorm(100,sd=sqrt(0.25))
```

Question 13c

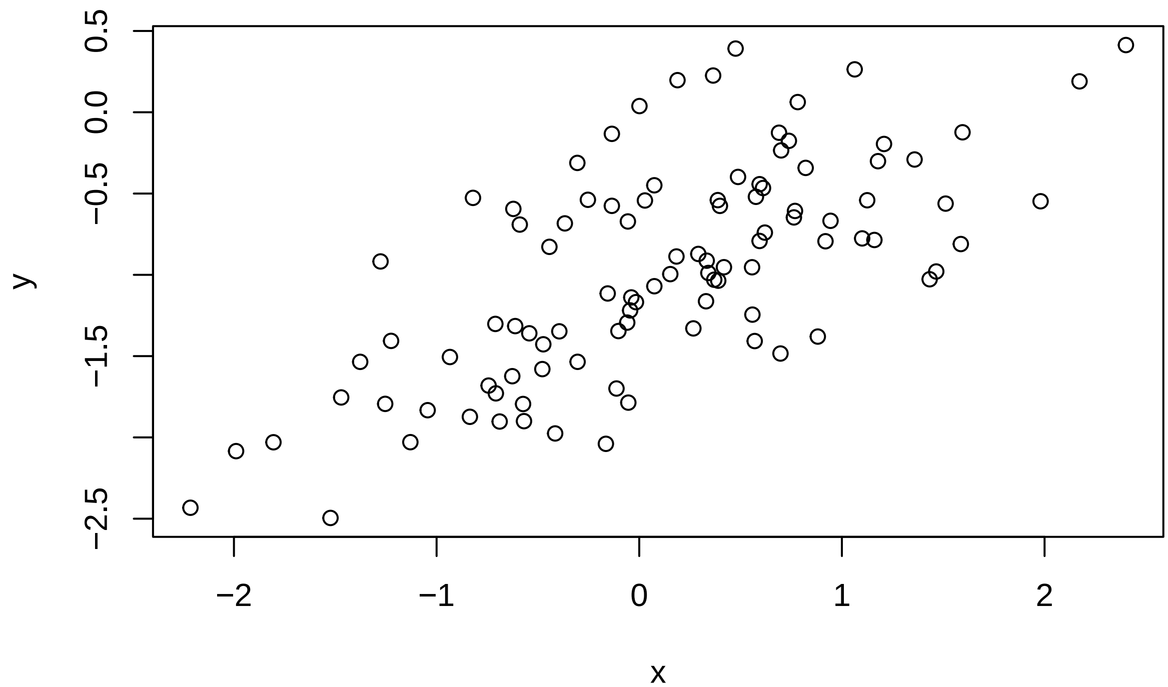
```
y<- -1+0.5*x+eps
print(length(y))
```

```
## [1] 100
```

B0=-1 B1=0.5

Question 13d

```
plot(x, y)
```



13d: The relationship between x and y looks to be mostly linear with slight noise introduced by the eps variable.

Question 13e

```
q13sim.fit<-lm(y~x)
summary(q13sim.fit)
```

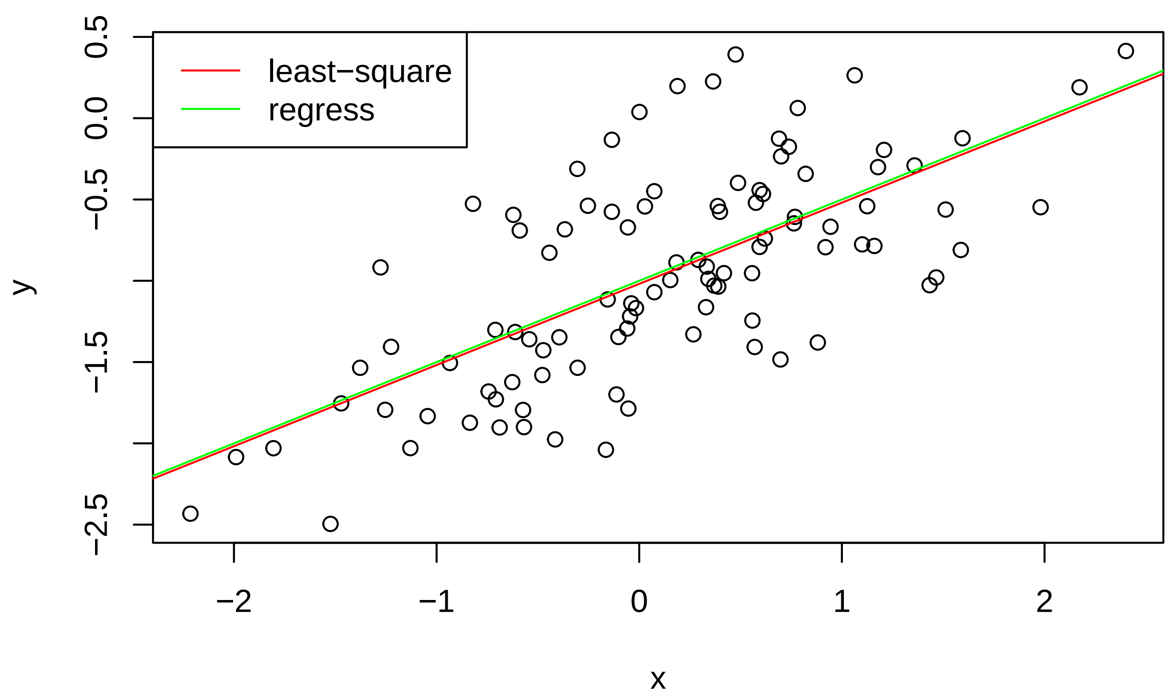
```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.93842 -0.30688 -0.06975  0.26970  1.17309
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.01885    0.04849 -21.010  < 2e-16 ***
## x            0.49947    0.05386   9.273 4.58e-15 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4814 on 98 degrees of freedom
## Multiple R-squared:  0.4674, Adjusted R-squared:  0.4619
## F-statistic: 85.99 on 1 and 98 DF,  p-value: 4.583e-15
```

13e: The hat values are pretty similar to the original values.

Question 13f

```r
plot(x,y)
abline(q13sim.fit, col="red")
abline(-1, 0.5, col="green")
legend("topleft", c("least-square","regress"), col=c("red", "green"), lty= c(1,1))
```



Question 13g

```r
q13gsim.fit<-lm(y~x+ I(x^2))
summary(q13gsim.fit)
```

```
##
## Call:
## lm(formula = y ~ x + I(x^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98252 -0.31270 -0.06441  0.29014  1.13500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.97164    0.05883 -16.517  < 2e-16 ***
## x            0.50858    0.05399   9.420  2.4e-15 ***
## I(x^2)      -0.05946    0.04238  -1.403    0.164
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.479 on 97 degrees of freedom
## Multiple R-squared:  0.4779, Adjusted R-squared:  0.4672
## F-statistic:  44.4 on 2 and 97 DF,  p-value: 2.038e-14
```

13g: There is not sufficient evidence that the quadratic term improves the model fit as its p-value is higher than 0.05.

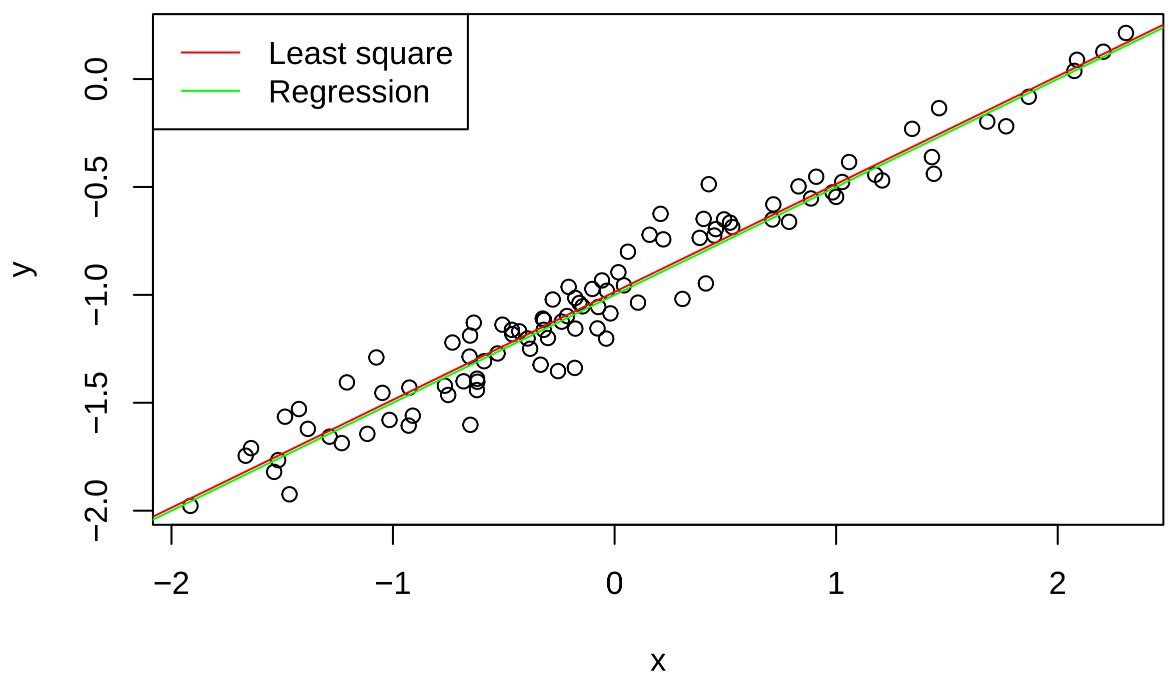
Question 13h

```r
set.seed(1)
eps <- rnorm(100, sd = 0.125)
x <- rnorm(100)
y <- -1 + 0.5 * x + eps
plot(x, y)
q13hsim.fit <- lm(y~x)
summary(q13hsim.fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29052 -0.07545  0.00067  0.07288  0.28664
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.98639    0.01129  -87.34   <2e-16 ***
```

```
## x            0.49988    0.01184   42.22   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1128 on 98 degrees of freedom
## Multiple R-squared:  0.9479, Adjusted R-squared:  0.9474
## F-statistic:  1782 on 1 and 98 DF,  p-value: < 2.2e-16
```

```r
abline(q13hsim.fit, col = "red")
abline(-1, 0.5, col = "green")
legend("topleft", c("Least square", "Regression"), col = c("red", "green"), lty = c(1, 1))
```



13h: Reduced the noise by decreasing the variance of the normal distribution used to generate the error term. The relationship is now mostly linear and has a much higher R^2.

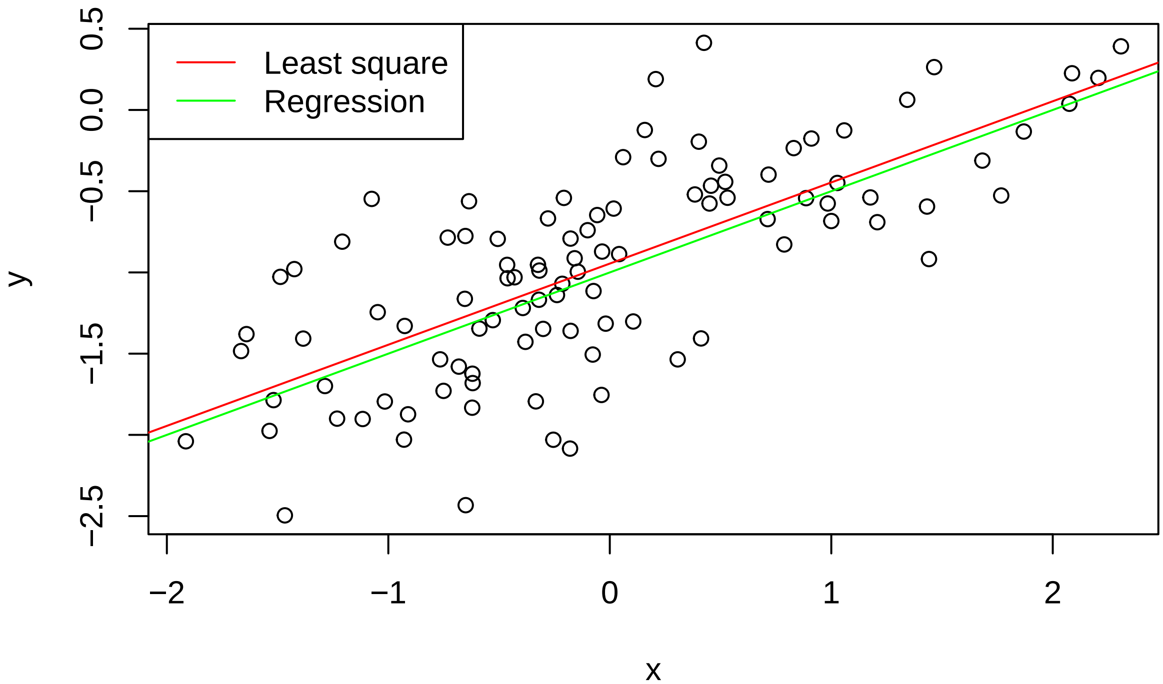
Question 13i

```r
set.seed(1)
eps<- rnorm(100, sd = 0.5)
x<- rnorm(100)
y<- -1 + 0.5 * x + eps
plot(x, y)
```

```r
q13isim.fit<- lm(y~x)
summary(q13isim.fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.16208 -0.30181  0.00268  0.29152  1.14658
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.94557    0.04517  -20.93   <2e-16 ***
## x            0.49953    0.04736   10.55   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4514 on 98 degrees of freedom
## Multiple R-squared:  0.5317, Adjusted R-squared:  0.5269
## F-statistic: 111.2 on 1 and 98 DF,  p-value: < 2.2e-16
```

```r
abline(q13isim.fit, col = "red")
abline(-1, 0.5, col = "green")
legend("topleft", c("Least square", "Regression"), col = c("red", "green"), lty = c(1, 1))
```



13i: We increased the noise by increasing the variance of the normal distribution used to generate the error term. We have a much lower R^2 and RSE.

Question 13j

```
confint(q13sim.fit)
```

```
##                  2.5 %     97.5 %
## (Intercept) -1.1150804 -0.9226122
## x            0.3925794  0.6063602
```

```
confint(q13hsim.fit)
```

```
##                 2.5 %     97.5 %
## (Intercept) -1.008805 -0.9639819
## x            0.476387  0.5233799
```

```
confint(q13isim.fit)
```

```
##                  2.5 %     97.5 %
## (Intercept) -1.0352203 -0.8559276
## x            0.4055479  0.5935197
```

13j: As noise increases, so does the interval. The inverse also seems to be true.

Question 15a

```r
data("Boston")
q15znsim.fit<-lm(crim~zn, data=Boston)
summary(q15znsim.fit)
```

```
##
## Call:
## lm(formula = crim ~ zn, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.429 -4.222 -2.620  1.250 84.523
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.45369    0.41722  10.675  < 2e-16 ***
## zn          -0.07393    0.01609  -4.594 5.51e-06 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.435 on 504 degrees of freedom
## Multiple R-squared:  0.04019,    Adjusted R-squared:  0.03828
## F-statistic:  21.1 on 1 and 504 DF,  p-value: 5.506e-06
```

```r
q15indus.fit<-lm(crim~indus, data=Boston)
summary(q15indus.fit)
```

```
##
## Call:
## lm(formula = crim ~ indus, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.972  -2.698  -0.736   0.712  81.813
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.06374    0.66723  -3.093  0.00209 **
## indus        0.50978    0.05102   9.991  < 2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.866 on 504 degrees of freedom
## Multiple R-squared:  0.1653, Adjusted R-squared:  0.1637
## F-statistic: 99.82 on 1 and 504 DF,  p-value: < 2.2e-16
```

```r
q15chas.fit<-lm(crim~chas, data=Boston)
summary(q15chas.fit)
```

```
##
## Call:
## lm(formula = crim ~ chas, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.738 -3.661 -3.435  0.018 85.232
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.7444     0.3961   9.453   <2e-16 ***
## chas         -1.8928     1.5061  -1.257    0.209
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.597 on 504 degrees of freedom
## Multiple R-squared:  0.003124,   Adjusted R-squared:  0.001146
## F-statistic: 1.579 on 1 and 504 DF,  p-value: 0.2094
```

```
q15nox.fit<-lm(crim~nox, data=Boston)
summary(q15nox.fit)
```

```
##
## Call:
## lm(formula = crim ~ nox, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.371  -2.738  -0.974   0.559  81.728
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -13.720      1.699  -8.073 5.08e-15 ***
## nox           31.249      2.999  10.419  < 2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.81 on 504 degrees of freedom
## Multiple R-squared:  0.1772, Adjusted R-squared:  0.1756
## F-statistic: 108.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
q15rm.fit<-lm(crim~rm, data=Boston)
summary(q15rm.fit)
```

```
##
## Call:
## lm(formula = crim ~ rm, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.604 -3.952 -2.654  0.989 87.197
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   20.482      3.365   6.088 2.27e-09 ***
## rm            -2.684      0.532  -5.045 6.35e-07 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.401 on 504 degrees of freedom
## Multiple R-squared:  0.04807,    Adjusted R-squared:  0.04618
## F-statistic: 25.45 on 1 and 504 DF,  p-value: 6.347e-07
```

```
q15age.fit<-lm(crim~age, data=Boston)
summary(q15age.fit)
```

```
##
## Call:
## lm(formula = crim ~ age, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.789 -4.257 -1.230  1.527 82.849
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.77791    0.94398  -4.002 7.22e-05 ***
## age          0.10779    0.01274   8.463 2.85e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.057 on 504 degrees of freedom
## Multiple R-squared:  0.1244, Adjusted R-squared:  0.1227
## F-statistic: 71.62 on 1 and 504 DF,  p-value: 2.855e-16
```

```
q15dis.fit<-lm(crim~dis, data=Boston)
summary(q15dis.fit)
```

```
##
## Call:
## lm(formula = crim ~ dis, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.708 -4.134 -1.527  1.516 81.674
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.4993     0.7304  13.006   <2e-16 ***
## dis          -1.5509     0.1683  -9.213   <2e-16 ***
## ---
## Signif. codes:
```

```
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.965 on 504 degrees of freedom
## Multiple R-squared:  0.1441, Adjusted R-squared:  0.1425
## F-statistic: 84.89 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
q15rad.fit<-lm(crim~rad, data=Boston)
summary(q15rad.fit)
```

```
##
## Call:
## lm(formula = crim ~ rad, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.164  -1.381  -0.141   0.660  76.433
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.28716    0.44348  -5.157 3.61e-07 ***
## rad          0.61791    0.03433  17.998  < 2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.718 on 504 degrees of freedom
## Multiple R-squared:  0.3913, Adjusted R-squared:   0.39
## F-statistic: 323.9 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
q15tax.fit<-lm(crim~tax, data=Boston)
summary(q15tax.fit)
```

```
##
## Call:
## lm(formula = crim ~ tax, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.513  -2.738  -0.194   1.065  77.696
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.528369   0.815809  -10.45   <2e-16 ***
## tax          0.029742   0.001847   16.10   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.997 on 504 degrees of freedom
## Multiple R-squared:  0.3396, Adjusted R-squared:  0.3383
## F-statistic: 259.2 on 1 and 504 DF,  p-value: < 2.2e-16
```

```r
q15ptratio.fit<-lm(crim~ptratio, data=Boston)
summary(q15ptratio.fit)
```

```
##
## Call:
## lm(formula = crim ~ ptratio, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.654 -3.985 -1.912  1.825 83.353
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.6469     3.1473  -5.607 3.40e-08 ***
## ptratio       1.1520     0.1694   6.801 2.94e-11 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.24 on 504 degrees of freedom
## Multiple R-squared:  0.08407,    Adjusted R-squared:  0.08225
## F-statistic: 46.26 on 1 and 504 DF,  p-value: 2.943e-11
```

```r
q15lstat.fit<-lm(crim~lstat, data=Boston)
summary(q15lstat.fit)
```

```
##
## Call:
## lm(formula = crim ~ lstat, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.925  -2.822  -0.664   1.079  82.862
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.33054    0.69376  -4.801 2.09e-06 ***
## lstat        0.54880    0.04776  11.491  < 2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.664 on 504 degrees of freedom
## Multiple R-squared:  0.2076, Adjusted R-squared:  0.206
## F-statistic:   132 on 1 and 504 DF,  p-value: < 2.2e-16
```

```r
q15medv.fit<-lm(crim~medv, data=Boston)
summary(q15medv.fit)
```

```
##
## Call:
```

```
## lm(formula = crim ~ medv, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.071 -4.022 -2.343  1.298 80.957
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.79654    0.93419   12.63   <2e-16 ***
## medv        -0.36316    0.03839   -9.46   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.934 on 504 degrees of freedom
## Multiple R-squared:  0.1508, Adjusted R-squared:  0.1491
## F-statistic: 89.49 on 1 and 504 DF,  p-value: < 2.2e-16
```

15a: Each predictor is statistically significant except for chas as it is the only predictor to have a p-value > 0.05.

Question 15b

```
q15bsim.fit<-lm(crim~.,data=Boston)
summary(q15bsim.fit)
```

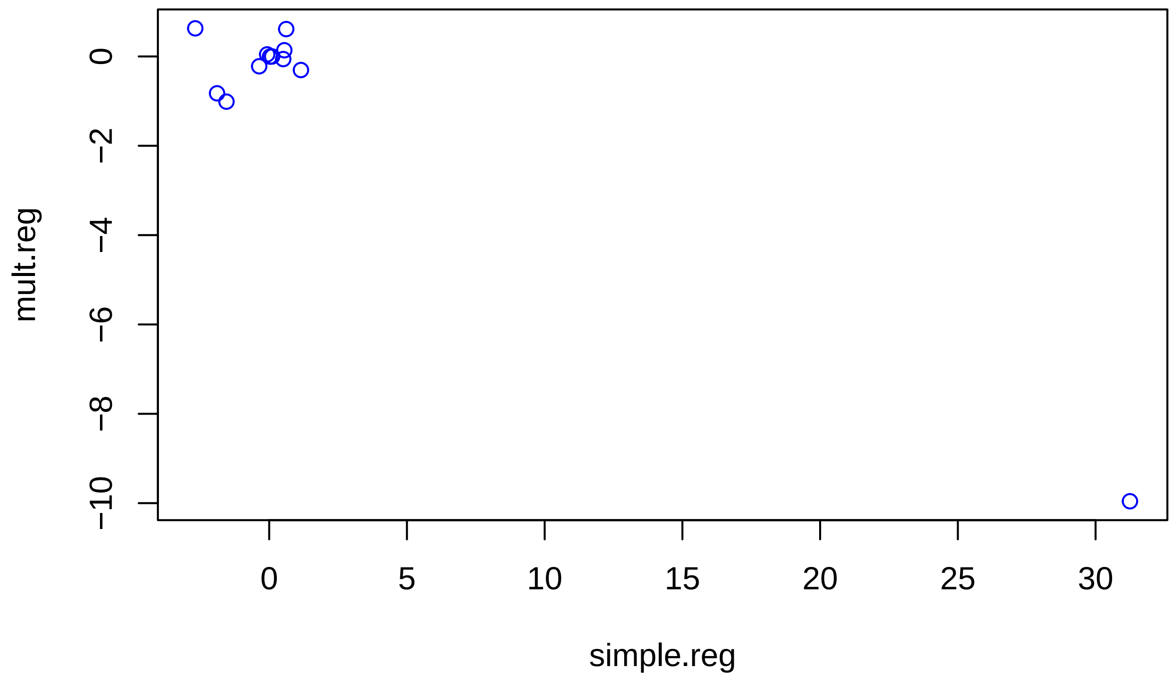
```
##
## Call:
## lm(formula = crim ~ ., data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.534 -2.248 -0.348  1.087 73.923
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.7783938  7.0818258   1.946 0.052271 .
## zn           0.0457100  0.0187903   2.433 0.015344 *
## indus       -0.0583501  0.0836351  -0.698 0.485709
## chas        -0.8253776  1.1833963  -0.697 0.485841
## nox         -9.9575865  5.2898242  -1.882 0.060370 .
## rm           0.6289107  0.6070924   1.036 0.300738
## age         -0.0008483  0.0179482  -0.047 0.962323
## dis         -1.0122467  0.2824676  -3.584 0.000373 ***
## rad          0.6124653  0.0875358   6.997 8.59e-12 ***
## tax         -0.0037756  0.0051723  -0.730 0.465757
## ptratio     -0.3040728  0.1863598  -1.632 0.103393
## lstat        0.1388006  0.0757213   1.833 0.067398 .
## medv        -0.2200564  0.0598240  -3.678 0.000261 ***
## ---
## Signif. codes:
```

```
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.46 on 493 degrees of freedom
## Multiple R-squared:  0.4493, Adjusted R-squared:  0.4359
## F-statistic: 33.52 on 12 and 493 DF,  p-value: < 2.2e-16
```

15b: From this, we can reject the null hypothesis for the following predictors: zn, dis, rad, and medv.

Question 15c

```
simple.reg<- vector("numeric",0)
simple.reg<- c(simple.reg, q15znsim.fit$coefficient[2])
simple.reg<- c(simple.reg, q15indus.fit$coefficient[2])
simple.reg<- c(simple.reg, q15chas.fit$coefficient[2])
simple.reg<- c(simple.reg, q15nox.fit$coefficient[2])
simple.reg<- c(simple.reg, q15rm.fit$coefficient[2])
simple.reg<- c(simple.reg, q15age.fit$coefficient[2])
simple.reg<- c(simple.reg, q15dis.fit$coefficient[2])
simple.reg<- c(simple.reg, q15rad.fit$coefficient[2])
simple.reg<- c(simple.reg, q15tax.fit$coefficient[2])
simple.reg<- c(simple.reg, q15ptratio.fit$coefficient[2])
simple.reg<- c(simple.reg, q15lstat.fit$coefficient[2])
simple.reg<- c(simple.reg, q15medv.fit$coefficient[2])
mult.reg<- vector("numeric", 0)
mult.reg<- c(mult.reg, q15bsim.fit$coefficients)
mult.reg<- mult.reg[-1]
plot(simple.reg, mult.reg, col = "blue")
```



15c: There is difference between the simple and multi regression coefficients due to that in the simple regresssion, the slope term represents the average effect of an increase in the predictor, while ignoring other predictors, but in the multi regress, the slope term represents the average effect of an increase in the predictor, while holding other predictors fixed. This difference allows for the multi regression to suggest no relationship between the response and some of the predictors while the simple regression implies the opposite since the correlations between the predictors show some strong relationships.

Question 15d

```r
zn2.fit<- lm(crim ~ poly(zn, 3), data=Boston)
summary(zn2.fit)
```

```
##
## Call:
## lm(formula = crim ~ poly(zn, 3), data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.821 -4.614 -1.294  0.473 84.130
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.6135     0.3722   9.709  < 2e-16 ***
## poly(zn, 3)1 -38.7498     8.3722  -4.628  4.7e-06 ***
```

```
## poly(zn, 3)2  23.9398     8.3722   2.859  0.00442 **
## poly(zn, 3)3 -10.0719     8.3722  -1.203  0.22954
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.372 on 502 degrees of freedom
## Multiple R-squared:  0.05824,    Adjusted R-squared:  0.05261
## F-statistic: 10.35 on 3 and 502 DF,  p-value: 1.281e-06
```

```r
indus2.fit <- lm(crim ~ poly(indus, 3), data=Boston)
summary(indus2.fit)
```

```
##
## Call:
## lm(formula = crim ~ poly(indus, 3), data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.278 -2.514  0.054  0.764 79.713
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)        3.614      0.330  10.950  < 2e-16 ***
## poly(indus, 3)1   78.591      7.423  10.587  < 2e-16 ***
## poly(indus, 3)2  -24.395      7.423  -3.286  0.00109 **
## poly(indus, 3)3  -54.130      7.423  -7.292  1.2e-12 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.423 on 502 degrees of freedom
## Multiple R-squared:  0.2597, Adjusted R-squared:  0.2552
## F-statistic: 58.69 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
nox2.fit <- lm(crim ~ poly(nox, 3), data=Boston)
summary(nox2.fit)
```

```
##
## Call:
## lm(formula = crim ~ poly(nox, 3), data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.110 -2.068 -0.255  0.739 78.302
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.6135     0.3216  11.237  < 2e-16 ***
## poly(nox, 3)1  81.3720     7.2336  11.249  < 2e-16 ***
## poly(nox, 3)2 -28.8286     7.2336  -3.985 7.74e-05 ***
## poly(nox, 3)3 -60.3619     7.2336  -8.345 6.96e-16 ***
## ---
```

```
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.234 on 502 degrees of freedom
## Multiple R-squared:  0.297,  Adjusted R-squared:  0.2928
## F-statistic: 70.69 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
rm2.fit <- lm(crim ~ poly(rm, 3), data=Boston)
summary(rm2.fit)
```

```
##
## Call:
## lm(formula = crim ~ poly(rm, 3), data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.485  -3.468  -2.221  -0.015  87.219
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.6135     0.3703   9.758  < 2e-16 ***
## poly(rm, 3)1 -42.3794     8.3297  -5.088 5.13e-07 ***
## poly(rm, 3)2  26.5768     8.3297   3.191  0.00151 **
## poly(rm, 3)3  -5.5103     8.3297  -0.662  0.50858
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.33 on 502 degrees of freedom
## Multiple R-squared:  0.06779,    Adjusted R-squared:  0.06222
## F-statistic: 12.17 on 3 and 502 DF,  p-value: 1.067e-07
```

```r
age2.fit <- lm(crim ~ poly(age, 3), data=Boston)
summary(age2.fit)
```

```
##
## Call:
## lm(formula = crim ~ poly(age, 3), data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.762 -2.673 -0.516  0.019 82.842
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.6135     0.3485  10.368  < 2e-16 ***
## poly(age, 3)1  68.1820     7.8397   8.697  < 2e-16 ***
## poly(age, 3)2  37.4845     7.8397   4.781 2.29e-06 ***
## poly(age, 3)3  21.3532     7.8397   2.724  0.00668 **
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 7.84 on 502 degrees of freedom
## Multiple R-squared:  0.1742, Adjusted R-squared:  0.1693
## F-statistic: 35.31 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
dis2.fit <- lm(crim ~ poly(dis, 3), data=Boston)
summary(dis2.fit)
```

```
##
## Call:
## lm(formula = crim ~ poly(dis, 3), data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.757  -2.588   0.031   1.267  76.378
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.6135     0.3259  11.087  < 2e-16 ***
## poly(dis, 3)1 -73.3886     7.3315 -10.010  < 2e-16 ***
## poly(dis, 3)2  56.3730     7.3315   7.689 7.87e-14 ***
## poly(dis, 3)3 -42.6219     7.3315  -5.814 1.09e-08 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.331 on 502 degrees of freedom
## Multiple R-squared:  0.2778, Adjusted R-squared:  0.2735
## F-statistic: 64.37 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
rad2.fit <- lm(crim ~ poly(rad, 3), data=Boston)
summary(rad2.fit)
```

```
##
## Call:
## lm(formula = crim ~ poly(rad, 3), data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.381  -0.412  -0.269   0.179  76.217
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.6135     0.2971  12.164  < 2e-16 ***
## poly(rad, 3)1 120.9074     6.6824  18.093  < 2e-16 ***
## poly(rad, 3)2  17.4923     6.6824   2.618  0.00912 **
## poly(rad, 3)3   4.6985     6.6824   0.703  0.48231
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.682 on 502 degrees of freedom
## Multiple R-squared:    0.4,  Adjusted R-squared:  0.3965
## F-statistic: 111.6 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
tax2.fit <- lm(crim ~ poly(tax, 3), data=Boston)
summary(tax2.fit)
```

```
##
## Call:
## lm(formula = crim ~ poly(tax, 3), data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.273  -1.389   0.046   0.536  76.950
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.6135     0.3047  11.860  < 2e-16 ***
## poly(tax, 3)1   112.6458     6.8537  16.436  < 2e-16 ***
## poly(tax, 3)2    32.0873     6.8537   4.682 3.67e-06 ***
## poly(tax, 3)3    -7.9968     6.8537  -1.167    0.244
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.854 on 502 degrees of freedom
## Multiple R-squared:  0.3689, Adjusted R-squared:  0.3651
## F-statistic:  97.8 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
ptratio2.fit <- lm(crim ~ poly(ptratio, 3), data=Boston)
summary(ptratio2.fit)
```

```
##
## Call:
## lm(formula = crim ~ poly(ptratio, 3), data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.833 -4.146 -1.655  1.408 82.697
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.614      0.361  10.008  < 2e-16 ***
## poly(ptratio, 3)1     56.045      8.122   6.901 1.57e-11 ***
## poly(ptratio, 3)2     24.775      8.122   3.050  0.00241 **
## poly(ptratio, 3)3    -22.280      8.122  -2.743  0.00630 **
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.122 on 502 degrees of freedom
## Multiple R-squared:  0.1138, Adjusted R-squared:  0.1085
## F-statistic: 21.48 on 3 and 502 DF,  p-value: 4.171e-13
```

```r
fit.lstat2 <- lm(crim ~ poly(lstat, 3), data=Boston)
summary(fit.lstat2)
```

```
##
## Call:
## lm(formula = crim ~ poly(lstat, 3), data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.234  -2.151  -0.486   0.066  83.353
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        3.6135     0.3392  10.654   <2e-16 ***
## poly(lstat, 3)1  88.0697     7.6294  11.543   <2e-16 ***
## poly(lstat, 3)2  15.8882     7.6294   2.082   0.0378 *
## poly(lstat, 3)3 -11.5740     7.6294  -1.517   0.1299
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.629 on 502 degrees of freedom
## Multiple R-squared:  0.2179, Adjusted R-squared:  0.2133
## F-statistic: 46.63 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
fit.medv2 <- lm(crim ~ poly(medv, 3), data=Boston)
summary(fit.medv2)
```

```
##
## Call:
## lm(formula = crim ~ poly(medv, 3), data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.427  -1.976  -0.437   0.439  73.655
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)        3.614      0.292  12.374  < 2e-16 ***
## poly(medv, 3)1  -75.058      6.569 -11.426  < 2e-16 ***
## poly(medv, 3)2   88.086      6.569  13.409  < 2e-16 ***
## poly(medv, 3)3  -48.033      6.569  -7.312 1.05e-12 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.569 on 502 degrees of freedom
## Multiple R-squared:  0.4202, Adjusted R-squared:  0.4167
## F-statistic: 121.3 on 3 and 502 DF,  p-value: < 2.2e-16
```

15d: For 'zn', 'rm', 'rad', 'tax', and 'lstat' as predictors, the p-values suggest that the cubic coefficient is not statistically significant. For 'indus', 'nox', 'age', 'dis', 'ptratio', and 'medv' as predictors, the p-values suggest the cubic coefficient is adequate.