STAT 388

Charles Hwang

Professor Matthews

STAT 388-001

20 September 2019

Section 3.7 Exercise 3.3

```
rm(list=ls())
# a) iii; given all other variables are held constant, the x5 = 10*x1*x3 term becomes negative for fema
# b) $137,100
x1 <- 4.00
x2 <- 110
x3 <- 1
x4 <- x1*x2
x5 <- x1*x3
ŷ = 50 + 20*x1 + 0.07*x2 + 35*x3 + 0.01*x4 - 10*x5
cat("$",ŷ*1000)
```

```
## $ 137100
```

```
# c) False; although the coefficient is very small, this does not tell us how much evidence there is of
```

Section 3.7 Exercise 3.8

```
rm(list=ls())
library(ISLR)                    # Exercise 3.8a
lm <- lm(mpg~horsepower,data=Auto)
summary(lm)
```

```
##
## 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
```

```
#   (i) Yes.
#   (ii) The relationship is fairly strong with r^2 = 0.6049.
#   (iii) The relationship is negative because the coefficient for horsepower is negative (-0.157845).
#   (iv) ~24.46705 miles-per-gallon, (97.98711, 98.01289)
hp <- 98
mpg <- 39.935861 - 0.157845*hp
cat(mpg,"miles-per-gallon")
```

```
## 24.46705 miles-per-gallon
```

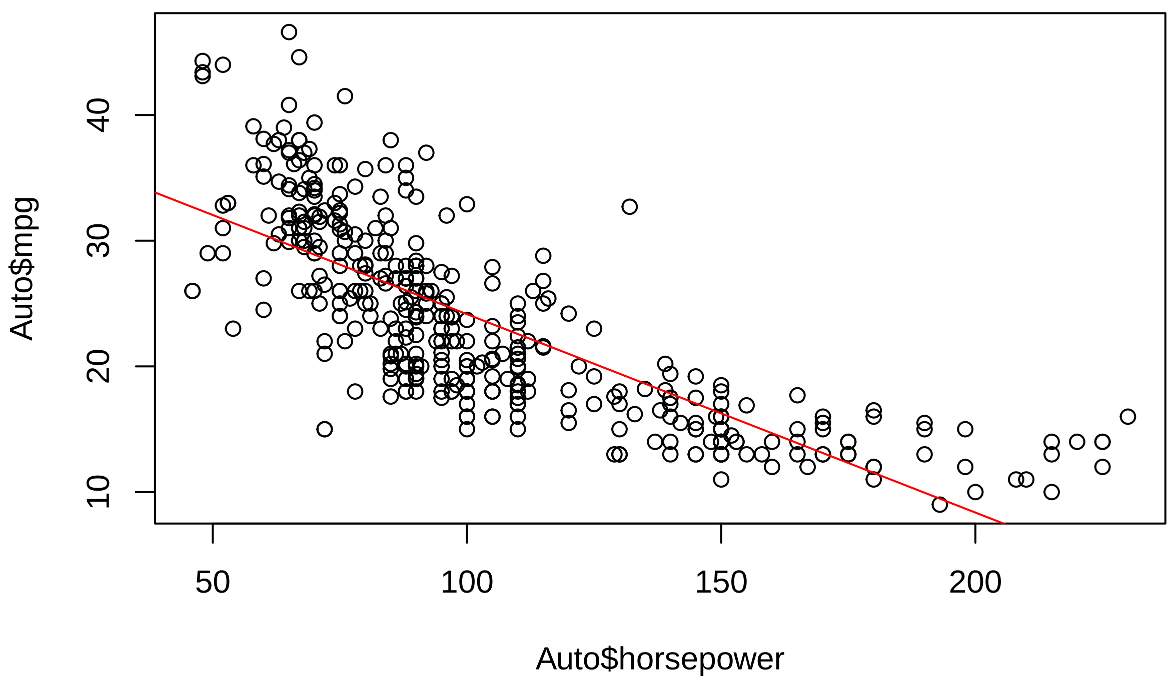
```
CI <- data.frame(horsepower=hp)
predict(lm,newdata=CI,interval="confidence")
```

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

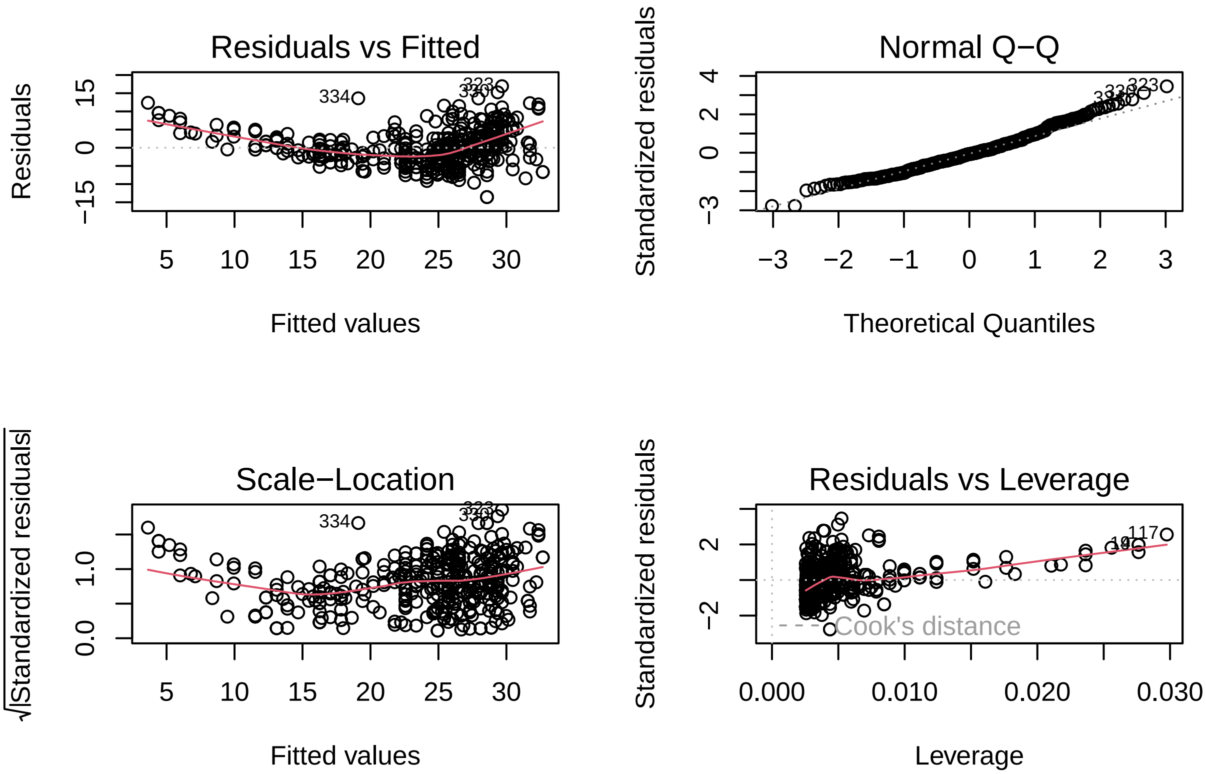
```
PI <- data.frame(horsepower=hp)
predict(lm,newdata=PI,interval="predict")
```

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

```
plot(Auto$mpg~Auto$horsepower) # Exercise 3.8b
abline(lm,col="red")
```




```
par(mfrow=c(2,2))             # Exercise 3.8c
plot(lm)
```




```
# There appears to be an upwards megaphone effect in the residual graph, potentially indicating that a
```

Section 3.7 Exercise 3.13

```
rm(list=ls())
set.seed(1)
x <- rnorm(100)                 # Exercise 3.13a
eps <- rnorm(100,0,sqrt(0.25)) # Exercise 3.13b
b0 <- -1                        # Exercise 3.13c
b1 <- 0.5
Y = b0 + b1*x + eps
length(Y)
```

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

```
plot(x,Y)                       # Exercise 3.13d
# The data form a roughly linear trend with a bulging effect in the middle.
lm <- lm(Y~x)                   # Exercise 3.13e
summary(lm)
```

```
##
## 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
```

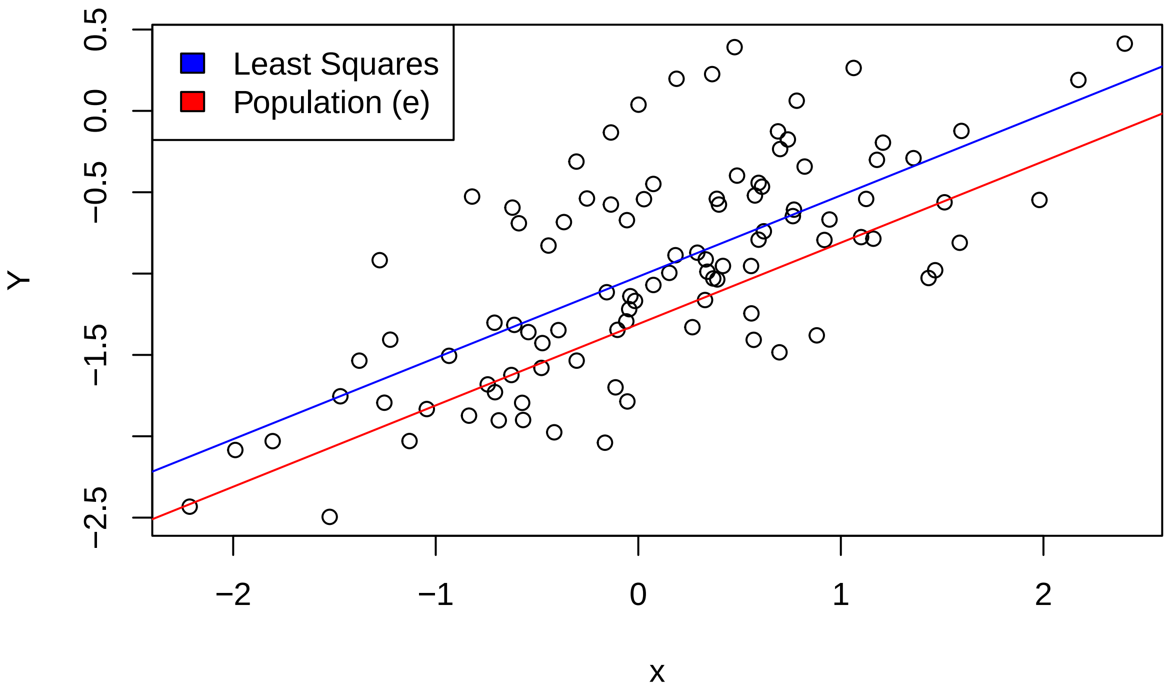
```
b0h <- lm$coefficients[1]
b1h <- lm$coefficients[2]
X <- data.frame(x=b1)
predict(lm,newdata=X,interval="predict")
```

```
##          fit       lwr       upr
## 1 -0.7691114 -1.730062 0.1918392
```

```
cat(b0h - b0,b1h - b1)
```

```
## -0.01884631 -0.0005301931
```

```
# b0-hat is 0.01884631 less than b0 and b1-hat is 0.0005301931 less than b1.
abline(lm,col="blue")            # Exercise 3.13f
abline(b0+eps,b1,col="red")
legend("topleft",legend=c("Least Squares","Population (e)"),fill=c("blue","red"))
```




```
x2 <- x^2                        # Exercise 3.13g
summary(lm(Y~x+x2))
```

```
##
## Call:
## lm(formula = Y ~ x + x2)
##
## 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 ***
## x2          -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
```

```
# No. The p-value for the quadratic term (0.164) is not significant at a = .05.
```

Section 3.7 Exercise 3.15

Exercise 3.15a

```
rm(list=ls())
library(MASS)
zn <- lm(crim~zn,data=Boston)
summary(zn)
```

```
##
## 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
```

```
indus <- lm(crim~indus,data=Boston)
summary(indus)
```

```
##
## 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
```

```
chas <- lm(crim~chas,data=Boston)
summary(chas)
```

```
##
## 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
```

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

```
##
## 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
```

```r
rm <- lm(crim~rm,data=Boston)
summary(rm)
```

```
##
## 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
```

```r
age <- lm(crim~age,data=Boston)
summary(age)
```

```
##
## 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
```

```r
dis <- lm(crim~dis,data=Boston)
summary(dis)
```

```
##
## 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
```

```r
rad <- lm(crim~rad,data=Boston)
summary(rad)
```

```
## 
## 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
```

```r
tax <- lm(crim~tax,data=Boston)
summary(tax)
```

```
## 
## 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
ptratio <- lm(crim~ptratio,data=Boston)
summary(ptratio)
```

```
##
## 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
black <- lm(crim~black,data=Boston)
summary(black)
```

```
##
## Call:
## lm(formula = crim ~ black, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.756  -2.299  -2.095  -1.296  86.822
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.553529   1.425903  11.609   <2e-16 ***
## black       -0.036280   0.003873  -9.367   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.946 on 504 degrees of freedom
## Multiple R-squared:  0.1483, Adjusted R-squared:  0.1466
## F-statistic: 87.74 on 1 and 504 DF,  p-value: < 2.2e-16
```

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

```
##
## Call:
## lm(formula = crim ~ lstat, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.925  -2.822  -0.664   1.079  82.862
##
```

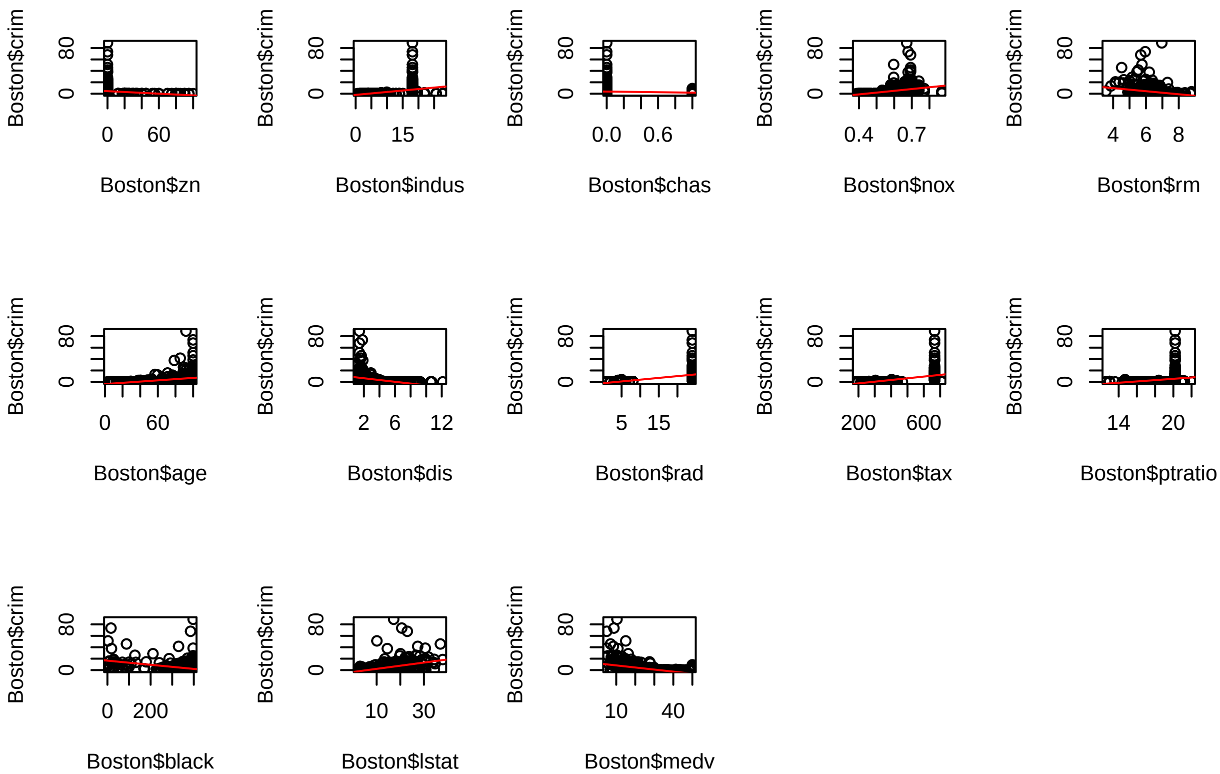
```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.33054    0.69376  -4.801 2.09e-06 ***
## lstat        0.54880    0.04776  11.491  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.664 on 504 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
```

```
medv <- lm(crim~medv,data=Boston)
summary(medv)
```

```
##
## 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
```

```
par(mfrow=c(3,5))
plot(Boston$crim~Boston$zn)
abline(zn,col="red")
plot(Boston$crim~Boston$indus)
abline(indus,col="red")
plot(Boston$crim~Boston$chas)
abline(chas,col="red")
plot(Boston$crim~Boston$nox)
abline(nox,col="red")
plot(Boston$crim~Boston$rm)
abline(rm,col="red")
plot(Boston$crim~Boston$age)
abline(age,col="red")
plot(Boston$crim~Boston$dis)
abline(dis,col="red")
plot(Boston$crim~Boston$rad)
abline(rad,col="red")
plot(Boston$crim~Boston$tax)
abline(tax,col="red")
plot(Boston$crim~Boston$ptratio)
abline(ptratio,col="red")
plot(Boston$crim~Boston$black)
```

```
abline(black,col="red")
plot(Boston$crim~Boston$lstat)
abline(lstat,col="red")
plot(Boston$crim~Boston$medv)
abline(medv,col="red")
# All of the models had a poor r^2, with the highest being the model for rad at r^2 = 0.39. The models
```



Exercise 3.15b

```
lm <- lm(crim~zn+indus+chas+nox+rm+age+dis+rad+tax+ptratio+black+lstat+medv,data=Boston)
summary(lm)
```

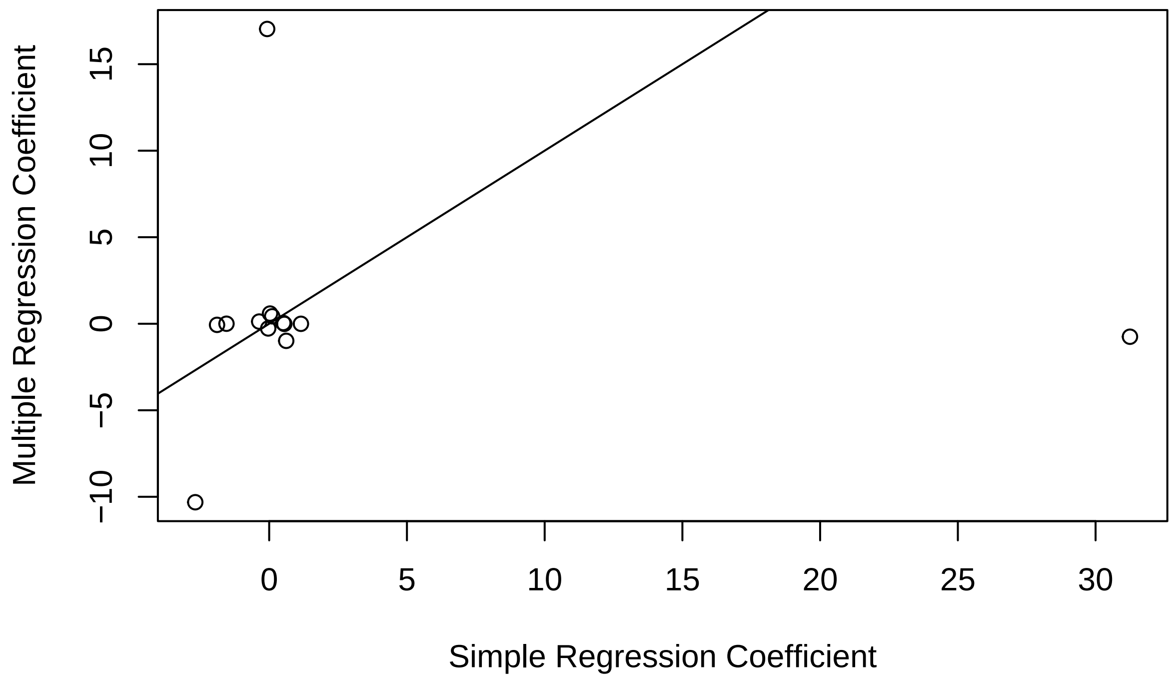
```
##
## Call:
## lm(formula = crim ~ zn + indus + chas + nox + rm + age + dis +
##     rad + tax + ptratio + black + lstat + medv, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.924 -2.120 -0.353  1.019 75.051
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.033228   7.234903   2.354 0.018949 *
## zn            0.044855   0.018734   2.394 0.017025 *
## indus        -0.063855   0.083407  -0.766 0.444294
## chas         -0.749134   1.180147  -0.635 0.525867
## nox         -10.313535   5.275536  -1.955 0.051152 .
## rm            0.430131   0.612830   0.702 0.483089
```

```
## age           0.001452   0.017925   0.081 0.935488
## dis          -0.987176   0.281817  -3.503 0.000502 ***
## rad           0.588209   0.088049   6.680 6.46e-11 ***
## tax          -0.003780   0.005156  -0.733 0.463793
## ptratio      -0.271081   0.186450  -1.454 0.146611
## black        -0.007538   0.003673  -2.052 0.040702 *
## lstat         0.126211   0.075725   1.667 0.096208 .
## medv         -0.198887   0.060516  -3.287 0.001087 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.439 on 492 degrees of freedom
## Multiple R-squared:  0.454,  Adjusted R-squared:  0.4396
## F-statistic: 31.47 on 13 and 492 DF,  p-value: < 2.2e-16
```

```
# There are many different variables with little effect on the overall model. There could be some multi
```

Exercise 3.15c

```
par(mfrow=c(1,1))
x <- c(zn$coefficients[2],indus$coefficients[2],chas$coefficients[2],nox$coefficients[2],rm$coefficient
y <- c(lm$coefficients[1],lm$coefficients[2],lm$coefficients[3],lm$coefficients[4],lm$coefficients[5],l
plot(x[1:13],y,xlab="Simple Regression Coefficient",ylab="Multiple Regression Coefficient")
abline(0,1) # Adding reference line
```



Exercise 3.15d

```
zn2 <- Boston$zn^2
zn3 <- Boston$zn^3
summary(lm(crim~zn+zn2+zn3,data=Boston))
```

```
##
```

```
## Call:
## lm(formula = crim ~ zn + zn2 + zn3, 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)  4.846e+00  4.330e-01  11.192  < 2e-16 ***
## zn          -3.322e-01  1.098e-01  -3.025  0.00261 **
## zn2          6.483e-03  3.861e-03   1.679  0.09375 .
## zn3         -3.776e-05  3.139e-05  -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
```

```
indus2 <- Boston$indus^2
indus3 <- Boston$indus^3
summary(lm(crim~indus+indus2+indus3,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ indus + indus2 + indus3, 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)  3.6625683  1.5739833   2.327   0.0204 *
## indus       -1.9652129  0.4819901  -4.077 5.30e-05 ***
## indus2       0.2519373  0.0393221   6.407 3.42e-10 ***
## indus3      -0.0069760  0.0009567  -7.292 1.20e-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
```

```
chas2 <- Boston$chas^2
chas3 <- Boston$chas^3
summary(lm(crim~chas+chas2+chas3,data=Boston)) # The chas variable does not have a quadratic or cubic t
```

```
##
## Call:
## lm(formula = crim ~ chas + chas2 + chas3, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -3.738 -3.661 -3.435  0.018 85.232
##
## Coefficients: (2 not defined because of singularities)
##             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
## chas2             NA         NA      NA       NA
## chas3             NA         NA      NA       NA
## ---
## 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
```

```r
nox2 <- Boston$nox^2
nox3 <- Boston$nox^3
summary(lm(crim~nox+nox2+nox3,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ nox + nox2 + nox3, 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)   233.09      33.64   6.928 1.31e-11 ***
## nox         -1279.37     170.40  -7.508 2.76e-13 ***
## nox2         2248.54     279.90   8.033 6.81e-15 ***
## nox3        -1245.70     149.28  -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 <- Boston$rm^2
rm3 <- Boston$rm^3
summary(lm(crim~rm+rm2+rm3,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ rm + rm2 + rm3, 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) 112.6246    64.5172   1.746   0.0815 .
```

```
## rm          -39.1501    31.3115  -1.250   0.2118
## rm2           4.5509     5.0099   0.908   0.3641
## rm3          -0.1745     0.2637  -0.662   0.5086
## ---
## 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
```

```
age2 <- Boston$age^2
age3 <- Boston$age^3
summary(lm(crim~age+age2+age3,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ age + age2 + age3, 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) -2.549e+00  2.769e+00  -0.920  0.35780
## age          2.737e-01  1.864e-01   1.468  0.14266
## age2        -7.230e-03  3.637e-03  -1.988  0.04738 *
## age3         5.745e-05  2.109e-05   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
```

```
dis2 <- Boston$dis^2
dis3 <- Boston$dis^3
summary(lm(crim~dis+dis2+dis3,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ dis + dis2 + dis3, 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)  30.0476     2.4459  12.285  < 2e-16 ***
## dis         -15.5543     1.7360  -8.960  < 2e-16 ***
## dis2          2.4521     0.3464   7.078 4.94e-12 ***
## dis3         -0.1186     0.0204  -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
```

```
rad2 <- Boston$rad^2
rad3 <- Boston$rad^3
summary(lm(crim~rad+rad2+rad3,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ rad + rad2 + rad3, 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) -0.605545   2.050108  -0.295    0.768
## rad          0.512736   1.043597   0.491    0.623
## rad2        -0.075177   0.148543  -0.506    0.613
## rad3         0.003209   0.004564   0.703    0.482
##
## 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
```

```
tax2 <- Boston$tax^2
tax3 <- Boston$tax^3
summary(lm(crim~tax+tax2+tax3,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ tax + tax2 + tax3, 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)  1.918e+01  1.180e+01   1.626    0.105
## tax         -1.533e-01  9.568e-02  -1.602    0.110
## tax2         3.608e-04  2.425e-04   1.488    0.137
## tax3        -2.204e-07  1.889e-07  -1.167    0.244
##
## 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
```

```
ptratio2 <- Boston$ptratio^2
ptratio3 <- Boston$ptratio^3
summary(lm(crim~ptratio+ptratio2+ptratio3,data=Boston))
```

```
##
```

```
## Call:
## lm(formula = crim ~ ptratio + ptratio2 + ptratio3, 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) 477.18405  156.79498   3.043  0.00246 **
## ptratio     -82.36054   27.64394  -2.979  0.00303 **
## ptratio2      4.63535    1.60832   2.882  0.00412 **
## ptratio3     -0.08476    0.03090  -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
```

```
black2 <- Boston$black^2
black3 <- Boston$black^3
summary(lm(crim~black+black2+black3,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ black + black2 + black3, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.096  -2.343  -2.128  -1.439  86.790
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.826e+01  2.305e+00   7.924  1.5e-14 ***
## black       -8.356e-02  5.633e-02  -1.483    0.139
## black2       2.137e-04  2.984e-04   0.716    0.474
## black3      -2.652e-07  4.364e-07  -0.608    0.544
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.955 on 502 degrees of freedom
## Multiple R-squared:  0.1498, Adjusted R-squared:  0.1448
## F-statistic: 29.49 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
lstat2 <- Boston$lstat^2
lstat3 <- Boston$lstat^3
summary(lm(crim~lstat+lstat2+lstat3,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ lstat + lstat2 + lstat3, 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)  1.2009656  2.0286452   0.592   0.5541
## lstat       -0.4490656  0.4648911  -0.966   0.3345
## lstat2       0.0557794  0.0301156   1.852   0.0646 .
## lstat3      -0.0008574  0.0005652  -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
```

```
medv2 <- Boston$medv^2
medv3 <- Boston$medv^3
summary(lm(crim~medv+medv2+medv3,data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ medv + medv2 + medv3, 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) 53.1655381  3.3563105  15.840  < 2e-16 ***
## medv        -5.0948305  0.4338321 -11.744  < 2e-16 ***
## medv2        0.1554965  0.0171904   9.046  < 2e-16 ***
## medv3       -0.0014901  0.0002038  -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
```

```
# Yes. The p-values for the quadratic terms are significant at a = .05 for the indus (0.000000000342),
```

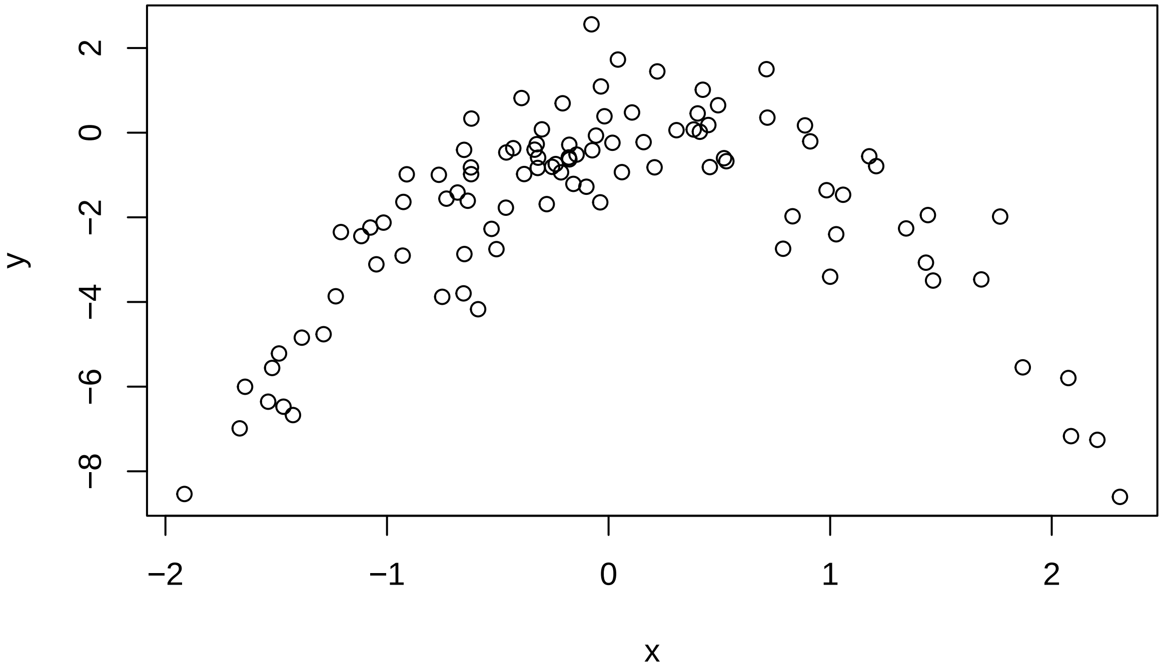
Section 5.4 Exercise 3

```
# a) The data are partitioned approximately equally into "k" groups, or "folds". One fold is treated as
# b) (i) k-fold cross-validation is more precise than the validation set approach because the partition
#    (ii) k-fold cross-validation takes less time than LOOCV, especially if the data set is large, beca
```

Section 5.4 Exercise 8

```
rm(list=ls())
set.seed(1)    # Exercise 5.8a
y = rnorm(100)
x = rnorm(100)
```

```r
  y = x - 2*x^2 + rnorm(100)
# n = 100, p = 1
# y = x - 2*x^2 + e
plot(x,y)      # Exercise 5.8b
```




```r
# There is a clear negative quadratic trend in the data.
set.seed(9)    # Exercise 5.8c
library(boot)
cv.glm(data.frame(x,y),glm(y~x,data=data.frame(x,y)))$delta                          # (i)
```

```
## [1] 5.890979 5.888812
```

```r
cv.glm(data.frame(x,y),glm(y~x+I(x^2),data=data.frame(x,y)))$delta                   # (ii)
```

```
## [1] 1.086596 1.086326
```

```r
cv.glm(data.frame(x,y),glm(y~x+I(x^2)+I(x^3),data=data.frame(x,y)))$delta            # (iii)
```

```
## [1] 1.102585 1.102227
```

```r
cv.glm(data.frame(x,y),glm(y~x+I(x^2)+I(x^3)+I(x^4),data=data.frame(x,y)))$delta # (iv)
```

```
## [1] 1.114772 1.114334
```

```r
# e) The quadratic model had the smallest LOOCV error. This is what was expected because the data clear
library(purrr) # Exercise 5.8f
map(list(glm(y~x,data=data.frame(x,y)),glm(y~x+I(x^2),data=data.frame(x,y)),glm(y~x+I(x^2)+I(x^3),data=
```

```
## [[1]]
##
## Call:
## glm(formula = y ~ x, data = data.frame(x, y))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -7.3469  -0.9275   0.8028   1.5608   4.3974
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.8185     0.2364  -7.692 1.14e-11 ***
## x             0.2430     0.2479   0.981    0.329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 5.580018)
##
##     Null deviance: 552.21  on 99  degrees of freedom
## Residual deviance: 546.84  on 98  degrees of freedom
## AIC: 459.69
##
## Number of Fisher Scoring iterations: 2
##
##
## [[2]]
##
## Call:
## glm(formula = y ~ x + I(x^2), data = data.frame(x, y))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.89884  -0.53765   0.04135   0.61490   2.73607
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.09544    0.13345  -0.715    0.476
## x            0.89961    0.11300   7.961 3.24e-12 ***
## I(x^2)      -1.86665    0.09151 -20.399  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.06575)
##
##     Null deviance: 552.21  on 99  degrees of freedom
## Residual deviance: 103.38  on 97  degrees of freedom
## AIC: 295.11
##
## Number of Fisher Scoring iterations: 2
##
##
## [[3]]
##
## Call:
## glm(formula = y ~ x + I(x^2) + I(x^3), data = data.frame(x, y))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.87250  -0.53881   0.02862   0.59383   2.74350
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -0.09865    0.13453  -0.733    0.465
## x            0.95551    0.22150   4.314  3.9e-05 ***
## I(x^2)      -1.85303    0.10296 -17.998  < 2e-16 ***
## I(x^3)      -0.02479    0.08435  -0.294    0.769
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.075883)
##
##     Null deviance: 552.21  on 99  degrees of freedom
## Residual deviance: 103.28  on 96  degrees of freedom
## AIC: 297.02
##
## Number of Fisher Scoring iterations: 2
##
##
## [[4]]
##
## Call:
## glm(formula = y ~ x + I(x^2) + I(x^3) + I(x^4), data = data.frame(x,
##     y))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8914  -0.5244   0.0749   0.5932   2.7796
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.13897    0.15973  -0.870 0.386455
## x            0.90980    0.24249   3.752 0.000302 ***
## I(x^2)      -1.72802    0.28379  -6.089  2.4e-08 ***
## I(x^3)       0.00715    0.10832   0.066 0.947510
## I(x^4)      -0.03807    0.08049  -0.473 0.637291
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.084654)
##
##     Null deviance: 552.21  on 99  degrees of freedom
## Residual deviance: 103.04  on 95  degrees of freedom
## AIC: 298.78
##
## Number of Fisher Scoring iterations: 2
```

```
# The linear and quadratic terms are significant for all models except the linear model where the linea
```

Problem 9

```
rm(list=ls())
p = 1
K = 2
pi1 = pi2 = 0.5
# 2x(m1 - m2) = m1^2 - m2^2
# x = (m1^2 - m2^2) / 2(m1 - m2)
```

```
# x = (m1 - m2)(m1 + m2) / 2(m1 - m2)
# x = (m1 + m2) / 2
```