HW2-MATH4322

anthonycastillo-ID:1670011

2022-09-14

Question 1a:

$\widehat{sales} = 2.938889 + 0.045765 * TV + 0.0188530 * radio - 0.001037 * newspaper + \epsilon$

Question 1b:

We will reject the null hypothesis for sales,TV and radio since the p-value is less than .05 and that means that there is significant evidence. The newspaper failed to reject the null hypothesis since the p-value is greater than .05>.86 and no significant evidence for the it.

Question 1c:

The newspaper predictor is not significant in predicating sales.

Question 2a:

```
TVRadioNewspaper.AIC <- 2*(3+1)+200*log(556.8/200)
TVRadio.AIC <- 2*(2+1)+200*log(556.9/200)
TV.AIC <- 2*(1+1)+200*log(2102.5/200)
TVRadioNewspaper.AIC
```

```
## [1] 212.7777
```

```
TVRadio.AIC
```

```
## [1] 210.8137
```

```
TV.AIC
```

```
## [1] 474.513
```

Question 2b:

```
TVRadioNewspaper.CP <- (556.8/2.8)+2*(3+1)-200
TVRadio.CP <- (556.9/2.8)+2*(2+1)-200
TV.CP <- (2102.5/2.8)+2*(1+1)-200
TVRadioNewspaper.CP
```

```
## [1] 6.857143
```

```
TVRadio.CP
```

```
## [1] 4.892857
```

```
TV.CP
```

```
## [1] 554.8929
```

Question 2c:

```
TVRadioNewspaper.AdjR2 <- 1-((556.8/196)/(5417.1/199))
TVRadio.AdjR2 <- 1-((556.9/197)/(5417.1/199))
TV.AdjR2 <- 1-((2102.5/198)/(5417.1/199))
TVRadioNewspaper.AdjR2
```

```
## [1] 0.8956411
```

```
TVRadio.AdjR2
```

```
## [1] 0.8961522
```

```
TV.AdjR2
```

```
## [1] 0.609917
```

Question 2d:

```
TVRadioNewspaper.RSE <- sqrt(556.8/(200-3-1))
TVRadio.RSE <- sqrt(556.9/(200-2-1))
TV.RSE <- sqrt(2102.5/(200-1-1))
TVRadioNewspaper.RSE
```

```
## [1] 1.685472
```

```
TVRadio.RSE
```

```
## [1] 1.68134
```

```
TV.RSE
```

```
## [1] 3.258633
```

Question 2e:
The second model best fits with only TV and Radio as predictors

Question 3a:
$\hat{y} = 50 + 20 * GPA + 0.07 * IQ + 35 * Gender + 0.01 * GPAxIQ - 10 * GPAxGender$
if X_3 = 1 then female and if male then X_3 = 0 so,
male : $\hat{y} = 50 + 20 * GPA + 0.07 * IQ + 0.01 * GPAxIQ$
female : $\hat{y} = 85 + 10 * GPA + 0.07 * IQ + 0.01 * GPAxIQ$
iii.is correct
since the GPA is higher in males than female that will make the males earn more on average than females

Question 3b:
IQ: 110 GPA:4.0

```
fem = (85 + (10*4.0) + (0.07*110) + (0.01*4.0*110))
fem
```

```
## [1] 137.1
```

Question 3c:
FALSE. can not tell if the predictors are significant without doing null hypothesis test and looking at the p-value

Question 4a:
TRUE
Question 4b:
TRUE
Question 4c:
FALSE
Question 4d:
FALSE
Question 4e:
FALSE

Question 5a:

```
library(ISLR2)
```

```
## Warning: package 'ISLR2' was built under R version 4.1.3
```

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

```
##
## 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. p-value is less than .05 meaning there is relationship

ii. has 60% variability somewhat strong relationship

iii. has a negative relationship

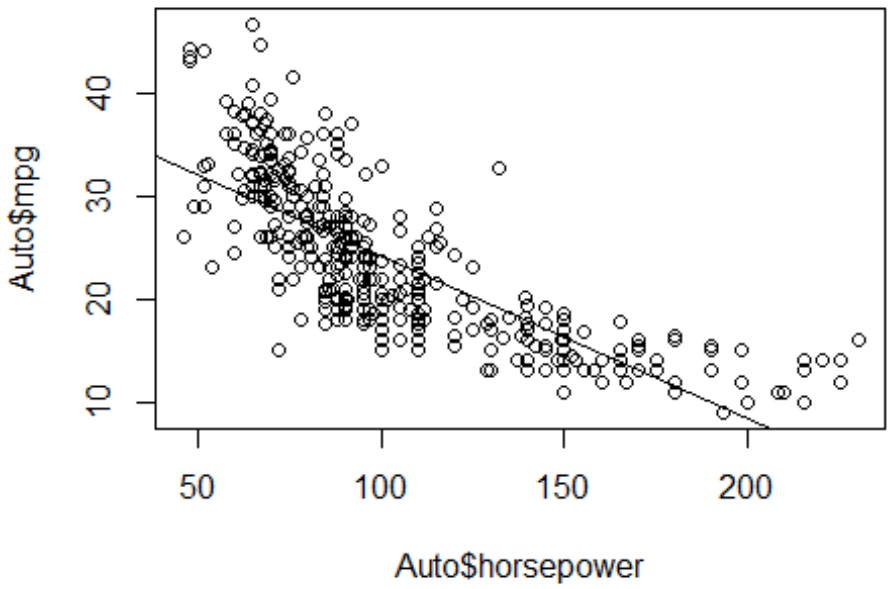
iv.

```
auto.lm = lm(mpg~horsepower, data = Auto)
predict(auto.lm,data.frame(horsepower=98), interval = "c")

##        fit      lwr      upr
## 1 24.46708 23.97308 24.96108

predict(auto.lm,data.frame(horsepower=98), interval = "p")

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

The predication interval has a bigger range compared to the confidence interval

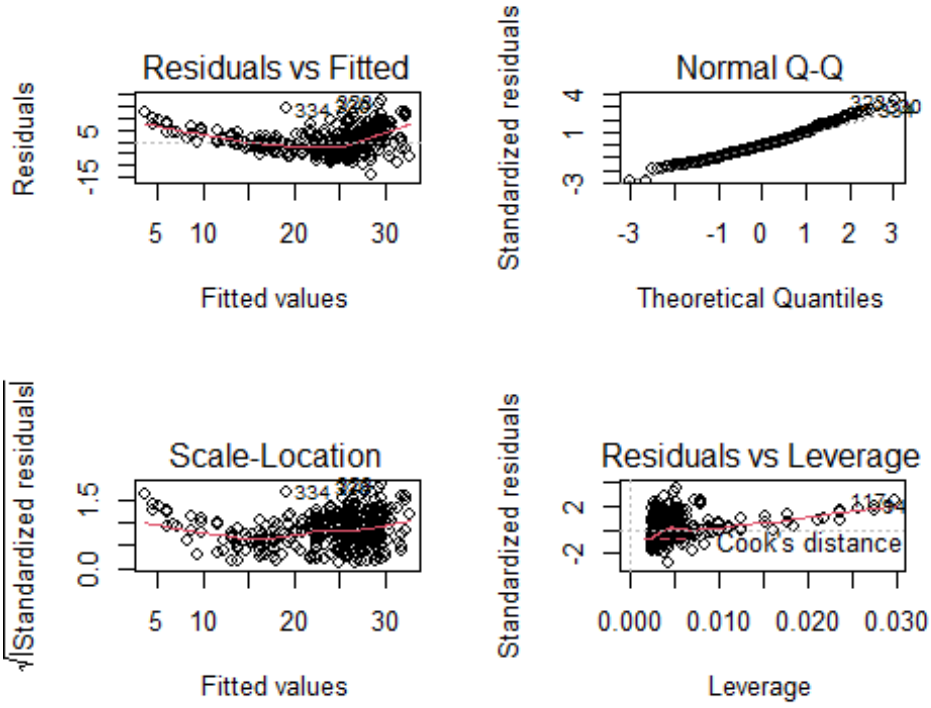
Question 5b:

```
plot(Auto$horsepower,Auto$mpg)
abline(auto.lm)
```



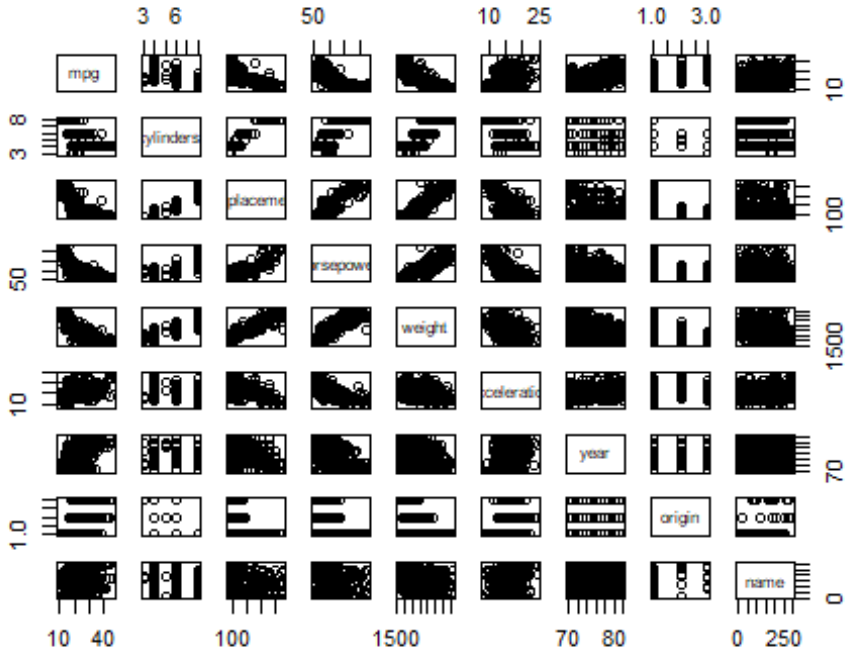
Question 5c:

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



Question 6a:

```
pairs(Auto)
```



Question 6b:

```
#colnames(Auto)
aa <- unlist(lapply(Auto, is.numeric))
data_aa <- Auto[ ,aa]
cor(data_aa)
```

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

Question 6c:

```
mpg.fit <- lm(mpg~
cylinders+displacement+horsepower+weight+acceleration+year+origin, data =
Auto)
summary(mpg.fit)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + origin, 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
```

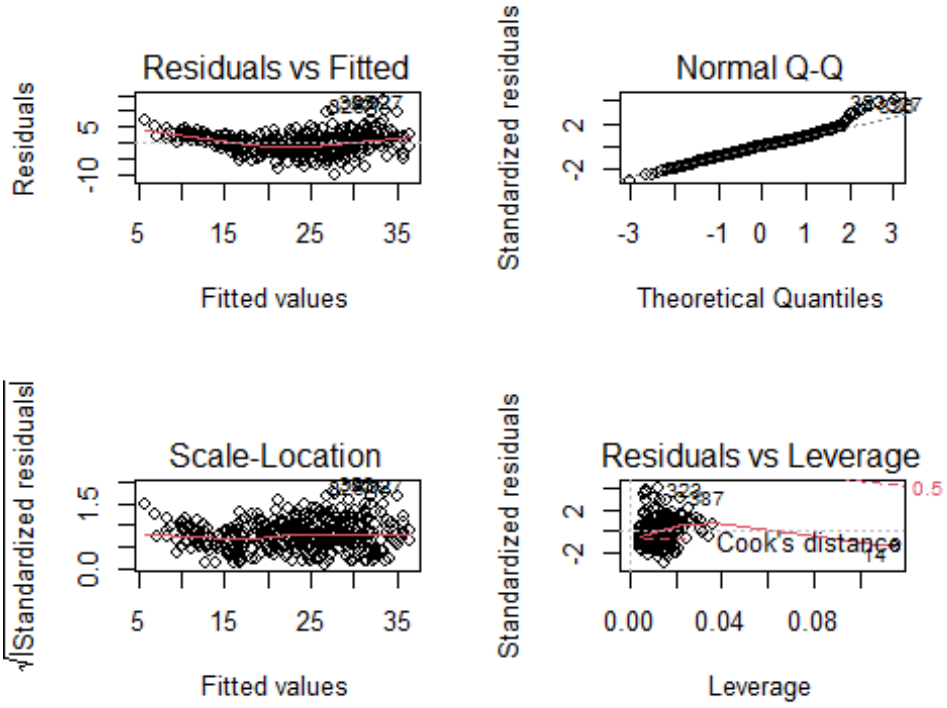
i. The p-value is less than 0.05 for mpg,displacement,weight,year, and origin meaning there is a relationship between them with the mpg predictor. The other variable are greater than .05 meaning that relationship is not that significant.

ii. displacement, weight, year and origin

iii. Their is a strong relationship with the mpg variable

Question 6d:

```
par(mfrow=c(2,2))
sigAuto.lm <- lm(mpg~ displacement+weight+year+origin, data = Auto)
plot(sigAuto.lm)
```



Question 6e:

```
aa.lm <- lm(mpg~ displacement*weight:year:origin, data = Auto)
summary(aa.lm)

##
## Call:
## lm(formula = mpg ~ displacement * weight:year:origin, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.071  -2.969  -0.530   2.297  18.828
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       3.476e+01  2.077e+00  16.736  < 2e-16 ***
## displacement                     -6.311e-02  1.358e-02  -4.647 4.62e-06 ***
## weight:year:origin                7.337e-07  6.011e-06   0.122    0.903
## displacement:weight:year:origin   1.174e-08  4.267e-08   0.275    0.783
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.638 on 388 degrees of freedom
## Multiple R-squared:  0.6495, Adjusted R-squared:  0.6468
## F-statistic: 239.7 on 3 and 388 DF,  p-value: < 2.2e-16
```

only displacement predictor is significant
the rest of the predictor are not significant
Question 6f:

```
logauto.lm <- lm(mpg~ log(displacement+weight+year+origin)^2, data = Auto)
summary(logauto.lm)
```

```
##
## Call:
## lm(formula = mpg ~ log(displacement + weight + year + origin)^2,
##     data = Auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.7598  -2.7104  -0.3992   2.0079  16.2746
##
## Coefficients:
##                                          Estimate Std. Error t value
Pr(>|t|)
## (Intercept)                              208.7827     5.9264   35.23
<2e-16
## log(displacement + weight + year + origin) -23.0375     0.7362  -31.29
<2e-16
##
## (Intercept)                              ***
## log(displacement + weight + year + origin) ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.171 on 390 degrees of freedom
## Multiple R-squared:  0.7152, Adjusted R-squared:  0.7144
## F-statistic: 979.3 on 1 and 390 DF,  p-value: < 2.2e-16
```

the log predictor is significant since p-value is less than 0.05

Question 7a:

```
#colnames(Boston)
summary(lm(crim~zn, data = Boston))
```

```
##
## 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
```

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

```
##
## 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
```

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

```
##
## 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
```

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

```
##
## 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
```

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

```
##
## 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
```

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

```
##
## 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
```

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

```
##
## 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
```

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

```
##
## 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
```

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

```
##
## 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
```

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

```
##
## 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
```

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

```
##
## 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
```

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

```
##
## 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
```

All of the predictors except chas have a p-value less than 0.05 meaning that all those predictors are significant
Question 7b:

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

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

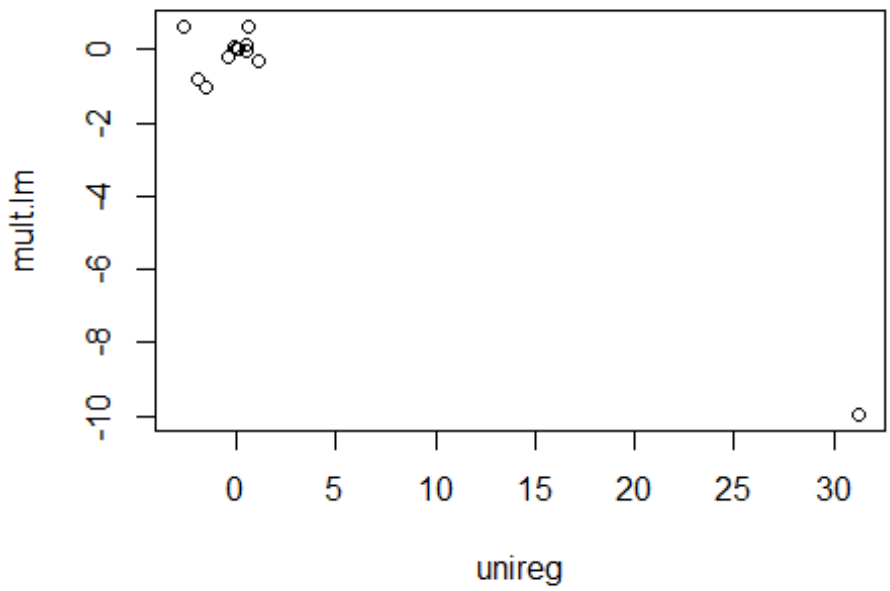
```
## lm(formula = crim ~ ., data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.534 -2.248 -0.348  1.087 73.923
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.7783938  7.0818258   1.946 0.052271 .
## zn           0.0457100  0.0187903   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
```

can reject null hypothesis for zn,dis,rad, and medv

Question 7c:

The multiple regression model has only 4 predictors that are significant compared to the simple linear regression model which has all predictors except “chas” that are significant.

```
unireg <- c(coefficients(lm(crim~zn, data = Boston))[2],
            coefficients(lm(crim~indus, data = Boston))[2],
            coefficients(lm(crim~chas, data = Boston))[2],
            coefficients(lm(crim~nox, data = Boston))[2],
            coefficients(lm(crim~rm, data = Boston))[2],
            coefficients(lm(crim~age, data = Boston))[2],
            coefficients(lm(crim~dis, data = Boston))[2],
            coefficients(lm(crim~rad, data = Boston))[2],
            coefficients(lm(crim~tax, data = Boston))[2],
            coefficients(lm(crim~ptratio, data = Boston))[2],
            coefficients(lm(crim~lstat, data = Boston))[2],
            coefficients(lm(crim~medv, data = Boston))[2]
            )
mult.lm <- (coefficients(crim.lm))[-1]
plot(unireg,mult.lm)
```




```
#coefficients(lm(crim~medv, data = Boston))
```

Question 7d:

```
summary(lm(crim~poly(zn,3), data = Boston))
```

```
##
## 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
summary(lm(crim~poly(indus,3), data = Boston))
```

```
##
## 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
#summary(lm(crim~poly(chas,3), data = Boston))
summary(lm(crim~poly(nox,3), data = Boston))
```

```
##
## 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
summary(lm(crim~poly(rm,3), data = Boston))
```

```
##
## 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

summary(lm(crim~poly(age,3), data = Boston))

##
## 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

summary(lm(crim~poly(dis,3), data = Boston))

##
## 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
```

```
summary(lm(crim~poly(rad,3), data = Boston))
```

```
##
## 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
```

```
summary(lm(crim~poly(tax,3), data = Boston))
```

```
##
## 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
```

```
summary(lm(crim~poly(ptratio,3), data = Boston))
```

```
##
## 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
```

```
summary(lm(crim~poly(lstat,3), data = Boston))
```

```
##
## 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
```

```
summary(lm(crim~poly(medv,3), data = Boston))
```

```
##
## 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
```

zn,rm,rad,tax,and lstat are not significant and do not fit the model
indus,nox,age,dis,ptratio, and medv are significant and fit the model

Question 8a:

```
set.seed(1)
x1 = runif(100)
x2 = .5* x1+rnorm(100)/10
y=2 + 2* x1 +0.3* x2 + rnorm(100)
```

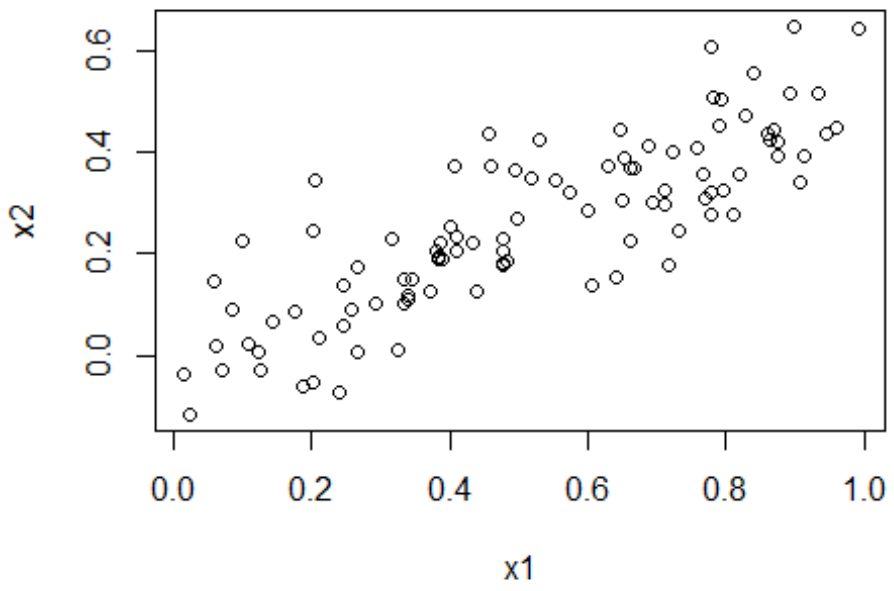
$$\hat{y} = 2 + 2 * X_1 + 0.3 * X_2 + \epsilon$$

coefficients are 2, 2, 0.3
Question 8b:

```
cor(x1,x2)
```

```
## [1] 0.8351212
```

```
plot(x1,x2)
```



Question 8c:

```
col.lm <- lm(y~x1+x2)
summary(col.lm)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8311 -0.7273 -0.0537  0.6338  2.3359
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1305     0.2319   9.188 7.61e-15 ***
## x1            1.4396     0.7212   1.996   0.0487 *
## x2            1.0097     1.1337   0.891   0.3754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.056 on 97 degrees of freedom
## Multiple R-squared:  0.2088, Adjusted R-squared:  0.1925
## F-statistic:  12.8 on 2 and 97 DF,  p-value: 1.164e-05
```

b0 = 2.1305 b1 = 1.4396 b2 = 1.009
We can reject both null hypothesis

Question 8d:

```
x1.lm <-lm(y~x1)
summary(x1.lm)
```

```
##
## Call:
## lm(formula = y ~ x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.89495 -0.66874 -0.07785  0.59221  2.45560
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1124     0.2307   9.155 8.27e-15 ***
## x1            1.9759     0.3963   4.986 2.66e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.055 on 98 degrees of freedom
## Multiple R-squared:  0.2024, Adjusted R-squared:  0.1942
## F-statistic: 24.86 on 1 and 98 DF,  p-value: 2.661e-06
```

has adj r^2 of 19% variability
reject null hypothesis p-value is less than 0.05
Question 8e:

```
x2.lm<-lm(y~x2)
summary(x2.lm)
```

```
##
## Call:
## lm(formula = y ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.62687 -0.75156 -0.03598  0.72383  2.44890
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3899     0.1949   12.26  < 2e-16 ***
## x2            2.8996     0.6330    4.58 1.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.072 on 98 degrees of freedom
## Multiple R-squared:  0.1763, Adjusted R-squared:  0.1679
## F-statistic: 20.98 on 1 and 98 DF,  p-value: 1.366e-05
```

has adj r^2 of 17% variability
reject null hypothesis p-value is less than 0.05
Question 8f:
Yes there is a contradiction in multiple regression predictors are not significant and in the simple linear regression both predictor are shown as significant.
Question 8g:

```
x1=c(x1,0.1)
x2=c(x2,0.8)
y=c(y,6)

new.lm <- lm(y~x1+x2)
newx1.lm <- lm(y~x1)
newx2.lm <- lm(y~x2)
summary(new.lm)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.73348 -0.69318 -0.05263  0.66385  2.30619
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2267     0.2314   9.624 7.91e-16 ***
## x1            0.5394     0.5922   0.911  0.36458
## x2            2.5146     0.8977   2.801  0.00614 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.075 on 98 degrees of freedom
## Multiple R-squared:  0.2188, Adjusted R-squared:  0.2029
## F-statistic: 13.72 on 2 and 98 DF,  p-value: 5.564e-06
```

```
summary(newx1.lm)
```

```
##
## Call:
## lm(formula = y ~ x1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8897 -0.6556 -0.0909  0.5682  3.5665
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2569     0.2390   9.445 1.78e-15 ***
```

```
## x1            1.7657     0.4124   4.282 4.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.111 on 99 degrees of freedom
## Multiple R-squared:  0.1562, Adjusted R-squared:  0.1477
## F-statistic: 18.33 on 1 and 99 DF,  p-value: 4.295e-05
```

```
summary(newx2.lm)
```

```
##
## Call:
## lm(formula = y ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.64729 -0.71021 -0.06899  0.72699  2.38074
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3451     0.1912  12.264  < 2e-16 ***
## x2            3.1190     0.6040   5.164 1.25e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.074 on 99 degrees of freedom
## Multiple R-squared:  0.2122, Adjusted R-squared:  0.2042
## F-statistic: 26.66 on 1 and 99 DF,  p-value: 1.253e-06
```

in the multiple regression model x2 is now significant while x1 is now insignificant
Question 9a:

```
set.seed(1)
n=100
X = rnorm(n)
error = rnorm(n)
```

Question 9b:

```
b0=66;b1=.3;b2=.03;b3=-3
Y=b0+b1*X+b2*X^2+b3*X^3+error
```

Question 9c:

```
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 4.1.3
```

```r
df <- data.frame(Y,X)
best.fit <- regsubsets(Y~poly(X,10), data = df)
fit.sum<-summary(best.fit)
fit.sum
```

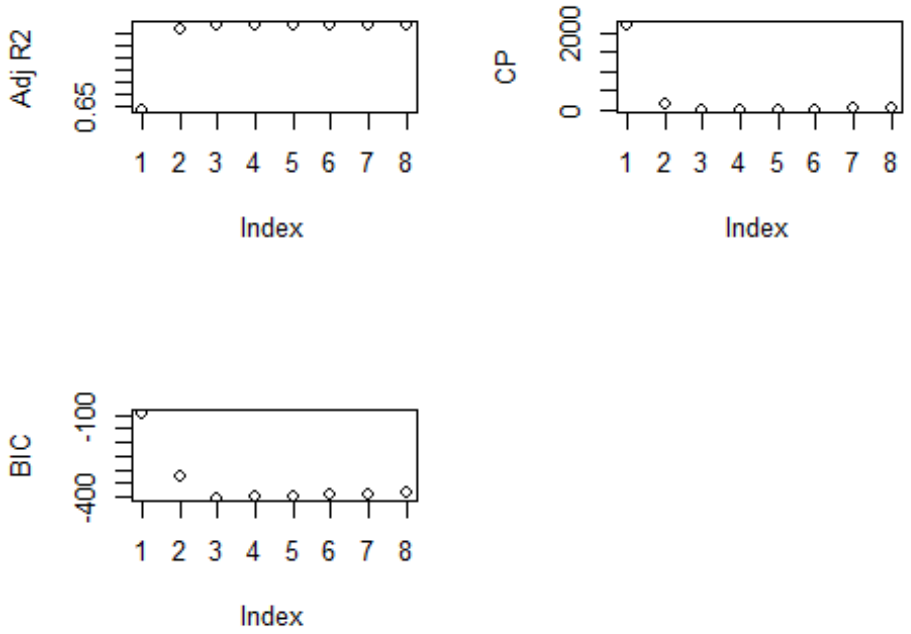
```
## Subset selection object
## Call: regsubsets.formula(Y ~ poly(X, 10), data = df)
## 10 Variables  (and intercept)
##               Forced in Forced out
## poly(X, 10)1      FALSE      FALSE
## poly(X, 10)2      FALSE      FALSE
## poly(X, 10)3      FALSE      FALSE
## poly(X, 10)4      FALSE      FALSE
## poly(X, 10)5      FALSE      FALSE
## poly(X, 10)6      FALSE      FALSE
## poly(X, 10)7      FALSE      FALSE
## poly(X, 10)8      FALSE      FALSE
## poly(X, 10)9      FALSE      FALSE
## poly(X, 10)10     FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##          poly(X, 10)1 poly(X, 10)2 poly(X, 10)3 poly(X, 10)4 poly(X, 10)5
## 1  ( 1 ) "*"          " "          " "          " "          " "
## 2  ( 1 ) "*"          " "          "*"          " "          " "
## 3  ( 1 ) "*"          "*"          "*"          " "          " "
## 4  ( 1 ) "*"          "*"          "*"          " "          "*"
## 5  ( 1 ) "*"          "*"          "*"          "*"          "*"
## 6  ( 1 ) "*"          "*"          "*"          "*"          "*"
## 7  ( 1 ) "*"          "*"          "*"          "*"          "*"
## 8  ( 1 ) "*"          "*"          "*"          "*"          "*"
##          poly(X, 10)6 poly(X, 10)7 poly(X, 10)8 poly(X, 10)9 poly(X, 10)10
## 1  ( 1 ) " "          " "          " "          " "          " "
## 2  ( 1 ) " "          " "          " "          " "          " "
## 3  ( 1 ) " "          " "          " "          " "          " "
## 4  ( 1 ) " "          " "          " "          " "          " "
## 5  ( 1 ) " "          " "          " "          " "          " "
## 6  ( 1 ) " "          " "          " "          " "          "*"
## 7  ( 1 ) " "          "*"          " "          " "          "*"
## 8  ( 1 ) " "          "*"          " "          "*"          "*"
```

```r
stat<- data.frame(
  Adj.R2 = which.max(fit.sum$adjr2),
  CP = which.min(fit.sum$cp),
  BIC = which.min(fit.sum$bic)
)
stat
```

```
##   Adj.R2 CP BIC
## 1      5  4   3
```

```
par(mfrow=c(2,2))
plot(fit.sum$adjr2, ylab="Adj R2")
abline(fit.sum)
plot(fit.sum$cp,ylab="CP")
plot(fit.sum$bic,ylab="BIC")

rstat = cbind(fit.sum$adjr2,fit.sum$cp,fit.sum$bic)
colnames(rstat) = c("adjr2","CP","BIC")
rstat
```

```
##          adjr2          CP        BIC
## [1,] 0.6339042 2223.225163  -92.29091
## [2,] 0.9668352  113.955971 -328.85195
## [3,] 0.9848228    2.185943 -403.45232
## [4,] 0.9850407    1.866261 -401.34086
## [5,] 0.9851569    2.193128 -398.57375
## [6,] 0.9851567    3.235128 -395.03643
## [7,] 0.9850147    5.119994 -390.56037
## [8,] 0.9848658    7.027330 -386.05923
```



The best model is only with the first predictor X
Question 9d:

```
step(lm(Y~poly(X,10)),direction  = "forward")
```

```
## Start:  AIC=4.64
## Y ~ poly(X, 10)
```

```
##
## Call:
## lm(formula = Y ~ poly(X, 10))
##
## Coefficients:
##   (Intercept)   poly(X, 10)1   poly(X, 10)2   poly(X, 10)3   poly(X, 10)4
##       65.3873       -62.0807       -10.3661       -44.6555         1.2571
##  poly(X, 10)5   poly(X, 10)6   poly(X, 10)7   poly(X, 10)8   poly(X, 10)9
##        1.4802         0.1190        -0.3298        -0.1079        -0.2958
## poly(X, 10)10
##       -0.9512
```

```
step(lm(Y~poly(X,10)),direction  = "backward")
```

```
## Start:  AIC=4.64
## Y ~ poly(X, 10)
##
##               Df Sum of Sq    RSS    AIC
## <none>                       84.1   4.64
## - poly(X, 10) 10    5960.5 6044.5 412.17

##
## Call:
## lm(formula = Y ~ poly(X, 10))
##
## Coefficients:
##   (Intercept)   poly(X, 10)1   poly(X, 10)2   poly(X, 10)3   poly(X, 10)4
##       65.3873       -62.0807       -10.3661       -44.6555         1.2571
##  poly(X, 10)5   poly(X, 10)6   poly(X, 10)7   poly(X, 10)8   poly(X, 10)9
##        1.4802         0.1190        -0.3298        -0.1079        -0.2958
## poly(X, 10)10
##       -0.9512
```