ANLY512_HW1

Hongyang Zheng

2019/2/4

Problem Ch.3 #3

The model is $\hat{salary} = 50 + 20X_1 + 0.07X_2 + 35X_3 + 0.01X4 - 10X5$

a)

iii is correct.

For female the model can be written as:

$$\hat{salary_{female}} = 50 + 20 * GPA + 0.07 * IQ + 35 + 0.01 * GPA * IQ - 10 * GPA$$

$$\hat{salary_{female}} = 85 + 10 * GPA + 0.01 * IQ * GPA + 0.07 * IQ$$

For male the model can be written as:

$$\hat{salary_{male}} = 50 + 20 * GPA + 0.07 * IQ + 0.01 * GPA * IQ$$

$$\hat{salary_{male}} = 50 + 20 * GPA + 0.01 * IQ * GPA + 0.07 * IQ$$

For a fixed value of IQ and GPA, if GPA is high enough such that $10 * GPA > 35$, then males earn more on average than females. Therefore, iii is the correct answer.

b)

When IQ=110 and GPA=4.0, $\hat{salary_{female}} = 85+10*4.0+0.01*110*4.0+0.07*110 = 85+40+4.4+7.7 = 137.1$

Therefore, the salary is 137.1*1000=137100 dollars

c)

The statement is false. Although the coefficient for the interactive term is small, if the p-value for this term is very small, then the interactive term is statistically significant.

Problem Ch.3 #8

a)

```r
# Load data
library(ISLR)
summary(Auto)
```

```
##       mpg          cylinders      displacement     horsepower
##  Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0
##  1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0
##  Median :22.75   Median :4.000   Median :151.0   Median : 93.5
##  Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5
##  3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0
```

```
##  Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0
##
##      weight      acceleration        year           origin
##  Min.   :1613   Min.   : 8.00   Min.   :70.00   Min.   :1.000
##  1st Qu.:2225   1st Qu.:13.78   1st Qu.:73.00   1st Qu.:1.000
##  Median :2804   Median :15.50   Median :76.00   Median :1.000
##  Mean   :2978   Mean   :15.54   Mean   :75.98   Mean   :1.577
##  3rd Qu.:3615   3rd Qu.:17.02   3rd Qu.:79.00   3rd Qu.:2.000
##  Max.   :5140   Max.   :24.80   Max.   :82.00   Max.   :3.000
##
##                 name
##  amc matador       :  5
##  ford pinto        :  5
##  toyota corolla    :  5
##  amc gremlin       :  4
##  amc hornet        :  4
##  chevrolet chevette:  4
##  (Other)           :365
```

```r
# Fit model
lm.fit1=lm(mpg~horsepower, data=Auto)
summary(lm.fit1)
```

```
##
## 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, there is a relationship between the predictor and the response, since the p-value for F-test is near 0. There is a strong evidence that the relationship between mpg and horsepower is statistically significant.

ii.

The R-squared is 0.6059, which means the predictor can explain about 60% of the variance in mpg, therefore, the relationship is strong.

iii.

Since the coefficient for horsepower is -0.157845, the relationship is negative.

iv.

```r
# Predicted mpg with horsepower of 98
mpg=39.935861+-0.157845*98

# Confidence and Prediction interval
predict(lm.fit1, data.frame(horsepower=c(98)), interval="confidence")
```

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

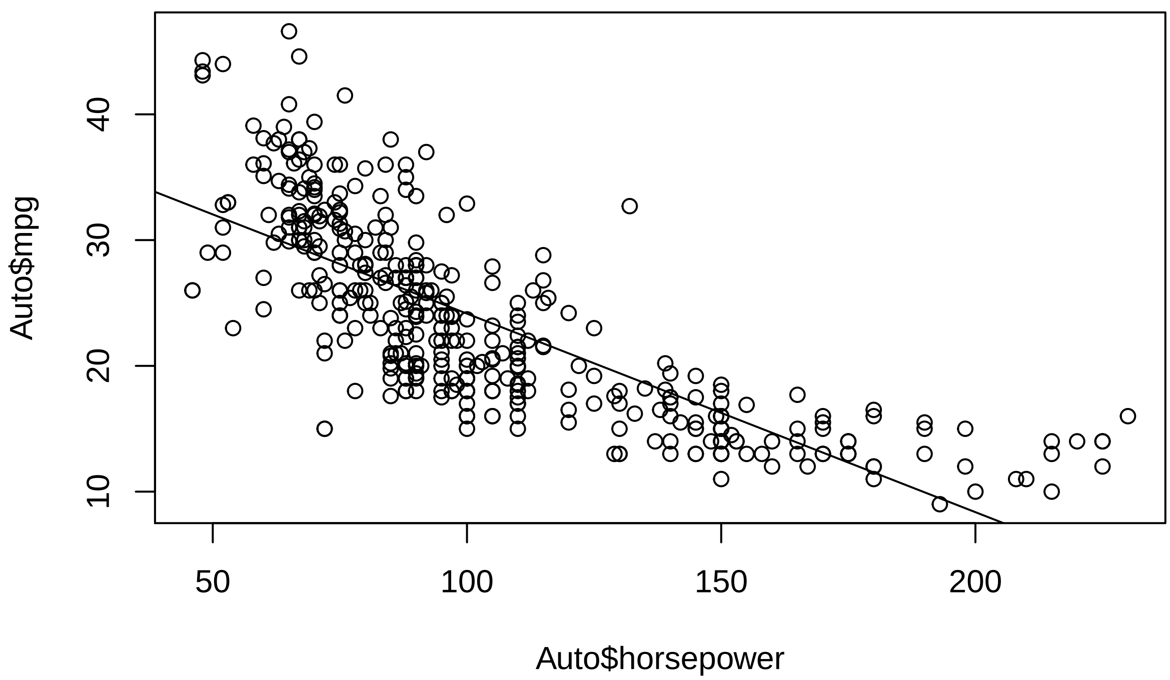
```r
predict(lm.fit1, data.frame(horsepower=c(98)), interval="prediction")
```

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

The predicted value is 24.467051; the confidence interval is $[23.97308, 24.96108]$ and the prediction interval is $[14.8094, 34.12476]$

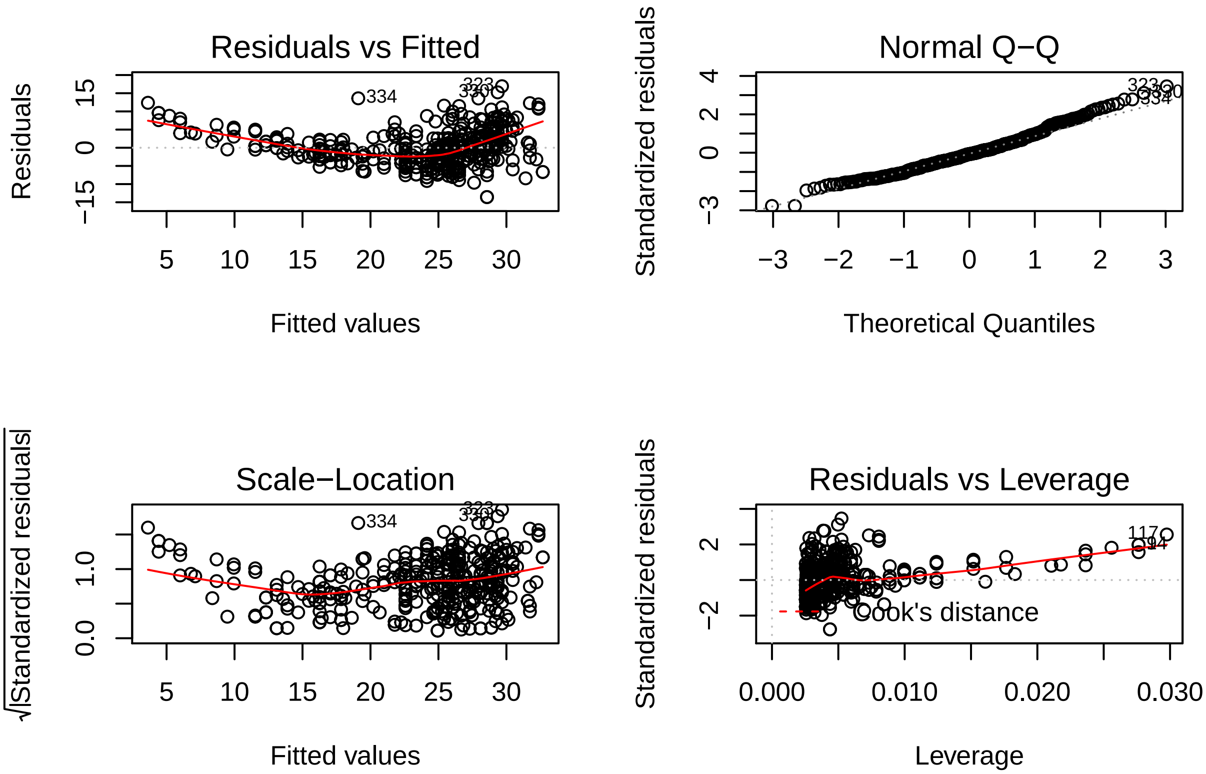
b)

```r
plot(Auto$mpg~Auto$horsepower)
abline(lm.fit1)
```



c)

```r
par(mfrow=c(2,2))
plot(lm.fit1)
```



From the first plot we can see that there is a non-linear relationship between the predictor and the response, since the points are not symmetric about the red line.

From the third plot we can see that the variance of residuals is not constant, since the points do not spread equally.

From the last plot we can see that some points have a standard residuals greater than 3, therefore, there are some outliers.

Problem Ch.3 #10abc

a)

```
lm.fit2=lm(Sales~Price+Urban+US, data=Carseats)
summary(lm.fit2)
```

```
##
## Call:
## lm(formula = Sales ~ Price + Urban + US, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9206 -1.6220 -0.0564  1.5786  7.0581
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.043469   0.651012  20.036  < 2e-16 ***
## Price       -0.054459   0.005242 -10.389  < 2e-16 ***
## UrbanYes    -0.021916   0.271650  -0.081    0.936
## USYes        1.200573   0.259042   4.635 4.86e-06 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.472 on 396 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2335
## F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16
```

b)

For price:

There is a relationship between Price and Sales since the p-value for Price is very small.

When a store is in an urban location of the US, if the price that company charges for car seats increases 1 unit, then the unit sales will decrease 54.459 units on average.

For Urban: There is no relationship between Urban and Sales since the p-value for Urban is very big.

For US: There is a relationship between US and Sales since the p-value for US is very small.

When a store is in an urban location and have a fixed price, the company in the US will sale 1200.573 more units than the company not in the US on average.

c)

$$\hat{Sales} = 13.04 - 0.05Price - 0.02UrbanYes + 1.20USYes$$

UrbanYes=1 when the company is in an urban location.

USYes=1 when the company is in the US.

Problem Ch.3 #15ab

a)

```r
library(MASS)
summary(Boston)
```

```
##       crim                zn             indus            chas
##  Min.   : 0.00632   Min.   :  0.00   Min.   : 0.46   Min.   :0.00000
##  1st Qu.: 0.08204   1st Qu.:  0.00   1st Qu.: 5.19   1st Qu.:0.00000
##  Median : 0.25651   Median :  0.00   Median : 9.69   Median :0.00000
##  Mean   : 3.61352   Mean   : 11.36   Mean   :11.14   Mean   :0.06917
##  3rd Qu.: 3.67708   3rd Qu.: 12.50   3rd Qu.:18.10   3rd Qu.:0.00000
##  Max.   :88.97620   Max.   :100.00   Max.   :27.74   Max.   :1.00000
##       nox               rm             age              dis
##  Min.   :0.3850   Min.   :3.561   Min.   :  2.90   Min.   : 1.130
##  1st Qu.:0.4490   1st Qu.:5.886   1st Qu.: 45.02   1st Qu.: 2.100
##  Median :0.5380   Median :6.208   Median : 77.50   Median : 3.207
##  Mean   :0.5547   Mean   :6.285   Mean   : 68.57   Mean   : 3.795
##  3rd Qu.:0.6240   3rd Qu.:6.623   3rd Qu.: 94.08   3rd Qu.: 5.188
##  Max.   :0.8710   Max.   :8.780   Max.   :100.00   Max.   :12.127
##       rad              tax           ptratio          black
##  Min.   : 1.000   Min.   :187.0   Min.   :12.60   Min.   :  0.32
##  1st Qu.: 4.000   1st Qu.:279.0   1st Qu.:17.40   1st Qu.:375.38
```

```
##  Median : 5.000   Median :330.0   Median :19.05   Median :391.44
##  Mean   : 9.549   Mean   :408.2   Mean   :18.46   Mean   :356.67
##  3rd Qu.:24.000   3rd Qu.:666.0   3rd Qu.:20.20   3rd Qu.:396.23
##  Max.   :24.000   Max.   :711.0   Max.   :22.00   Max.   :396.90
##      lstat            medv
##  Min.   : 1.73   Min.   : 5.00
##  1st Qu.: 6.95   1st Qu.:17.02
##  Median :11.36   Median :21.20
##  Mean   :12.65   Mean   :22.53
##  3rd Qu.:16.95   3rd Qu.:25.00
##  Max.   :37.97   Max.   :50.00
```

```
lm.model1=lm(crim~zn,data=Boston)
summary(lm.model1)
```

```
##
## 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
```

```
lm.model2=lm(crim~indus,data=Boston)
summary(lm.model2)
```

```
##
## 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
lm.model3=lm(crim~chas,data=Boston)
summary(lm.model3)
```

```
##
## 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
```

```r
lm.model4=lm(crim~nox,data=Boston)
summary(lm.model4)
```

```
##
## 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
lm.model5=lm(crim~rm,data=Boston)
summary(lm.model5)
```

```
##
## 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
lm.model6=lm(crim~age,data=Boston)
summary(lm.model6)
```

```
## 
## 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
lm.model7=lm(crim~dis,data=Boston)
summary(lm.model7)
```

```
## 
## 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
lm.model8=lm(crim~rad,data=Boston)
summary(lm.model8)
```

```
##
## 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
lm.model9=lm(crim~tax,data=Boston)
summary(lm.model9)
```

```
##
## 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
lm.model10=lm(crim~ptratio,data=Boston)
summary(lm.model10)
```

```
##
## 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
lm.model11=lm(crim~black,data=Boston)
summary(lm.model11)
```

```
##
## 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
lm.model12=lm(crim~lstat,data=Boston)
summary(lm.model12)
```

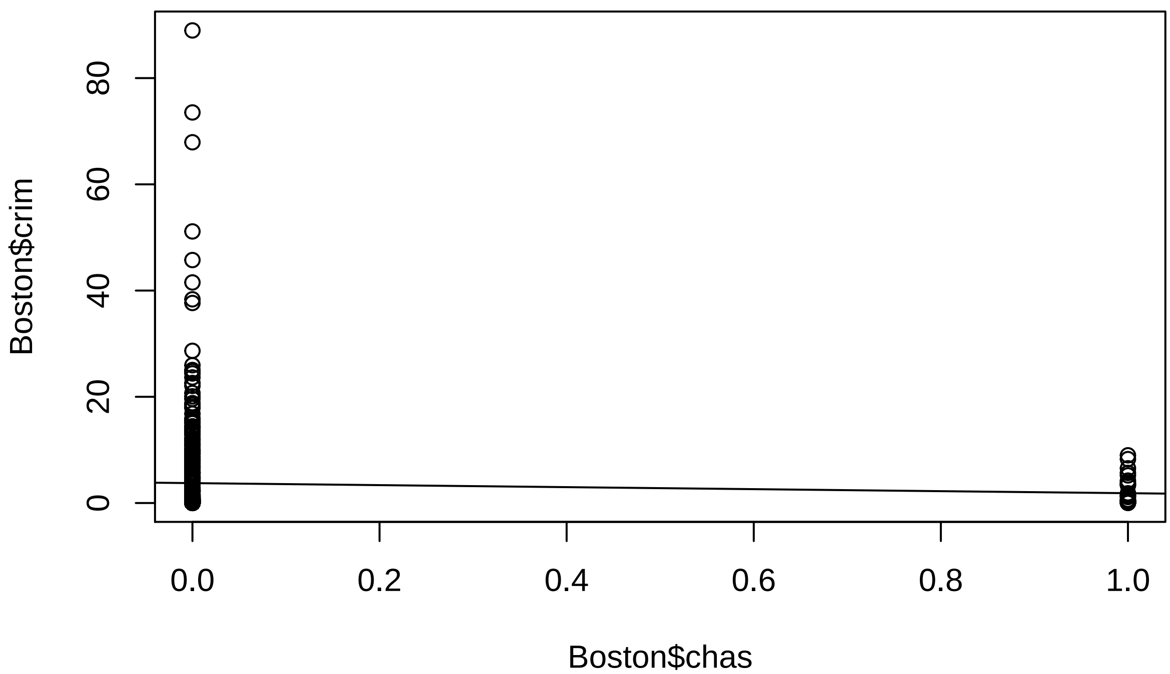
```
##
## Call:
## lm(formula = crim ~ lstat, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.925  -2.822  -0.664   1.079  82.862
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.33054    0.69376  -4.801 2.09e-06 ***
## lstat        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
lm.model13=lm(crim~medv,data=Boston)
summary(lm.model13)
```

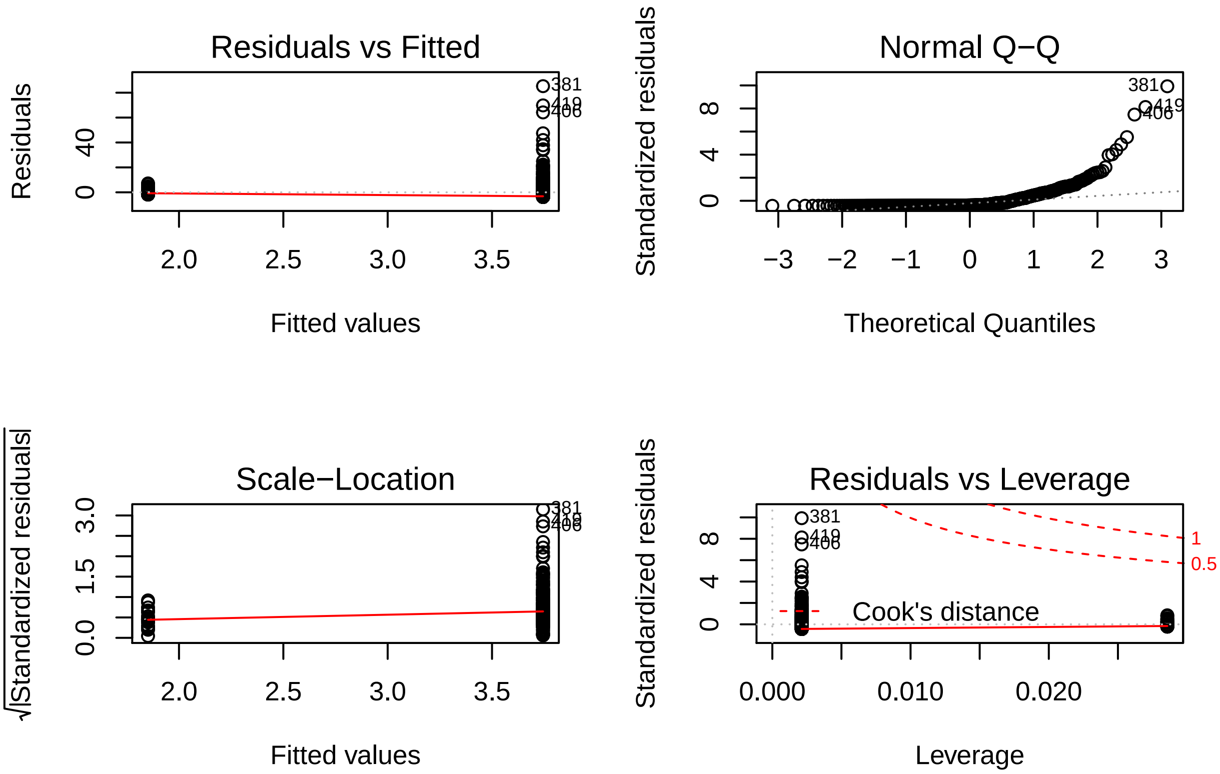
```
##
## Call:
## lm(formula = crim ~ medv, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.071 -4.022 -2.343  1.298 80.957
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.79654    0.93419   12.63   <2e-16 ***
## medv        -0.36316    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 predictors except 'chas' have a statistically significant coefficient.

```r
plot(Boston$chas, Boston$crim)
abline(lm.model3)
```




```r
par(mfrow=c(2,2))
plot(lm.model3)
```



b)

```
lm.model=lm(crim~.,data=Boston)
summary(lm.model)
```

```
##
## Call:
## lm(formula = crim ~ ., 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 is a relationship between predictors and response since the p-value for F-test is near 0. The R-squared is 0.454, which is okay. For predictors: zn, nox, dis, rad, black, lstat and medv, we can reject the null hypothesis that $\beta_j = 0$, since the p-value for those coefficients is smaller than 0.10.

Problem Xtra #10

```r
# Fit the model
lm.cars=lm(dist~speed, data=cars)

# Influence.lm
influence.cars=influence(lm.cars)

# The hat contains the value of the diagonal of the 'hat' matrix, which is the leverage
leverage.cars=sort(influence.cars$hat, decreasing=TRUE)
leverage.cars
```

```
##          1          2         50         46         47         48
## 0.11486131 0.11486131 0.08727007 0.07398540 0.07398540 0.07398540
##         49          3          4         45          5         44
## 0.07398540 0.07150365 0.07150365 0.06216058 0.05997080 0.05179562
##          6          7          8          9         39         40
## 0.04989781 0.04128467 0.04128467 0.04128467 0.03544526 0.03544526
##         41         42         43         10         11         36
## 0.03544526 0.03544526 0.03544526 0.03413139 0.03413139 0.02945985
##         37         38         12         13         14         15
## 0.02945985 0.02945985 0.02843796 0.02843796 0.02843796 0.02843796
##         32         33         34         35         16         17
## 0.02493431 0.02493431 0.02493431 0.02493431 0.02420438 0.02420438
##         18         19         29         30         31         20
## 0.02420438 0.02420438 0.02186861 0.02186861 0.02186861 0.02143066
##         21         22         23         27         28         24
## 0.02143066 0.02143066 0.02143066 0.02026277 0.02026277 0.02011679
##         25         26
## 0.02011679 0.02011679
```

```r
# The first three observations
leverage.cars[1:3]
```

```
##          1          2         50
## 0.11486131 0.11486131 0.08727007
```

```r
# Calculate standardized residuals
sresiduals.cars=sort(abs(influence.cars$wt.res/influence.cars$sigma), decreasing=TRUE)
sresiduals.cars
```

```
##         49         23         35         39         34         22
## 3.06490754 2.99026274 2.07215483 1.94693867 1.50237089 1.48356341
```

```
##          24         36         45         29         48         12
## 1.40661184 1.38837813 1.23467741 1.12651129 1.05565798 1.01562131
##          25         47         27         40         26         19
## 1.00204257 0.98898543 0.86521328 0.84750475 0.81607332 0.80721227
##           9          4          2         20         32         37
## 0.79398850 0.78084369 0.76778134 0.74269591 0.72500898 0.72060459
##          38         13         30         41         10          6
## 0.70272387 0.62090343 0.59887714 0.58571174 0.56023405 0.50409283
##          16         46          3         14         28         42
## 0.48651992 0.43839671 0.38332530 0.36145331 0.34398306 0.32652988
##          50          8          1          7         44         43
## 0.27490293 0.27399677 0.24785853 0.24110965 0.18884362 0.18865050
##          33         11          5         15         21         31
## 0.17994234 0.14947365 0.13641906 0.10358798 0.09488740 0.04685031
##          17         18
## 0.02945409 0.02945409
```

```r
# The first three observations
sresiduals.cars[1:3]
```

```
##       49       23       35
## 3.064908 2.990263 2.072155
```

The three observations with largest standardized residuals are 49, 23, 35, and their corresponding leverage are 0.07398540, 0.02143066, 0.02493431.

The three observations with largest leverage are 1, 2, 50, and their corresponding standardized residuals (in magnitude) are 0.24785853, 0.76778134, 0.27490293.

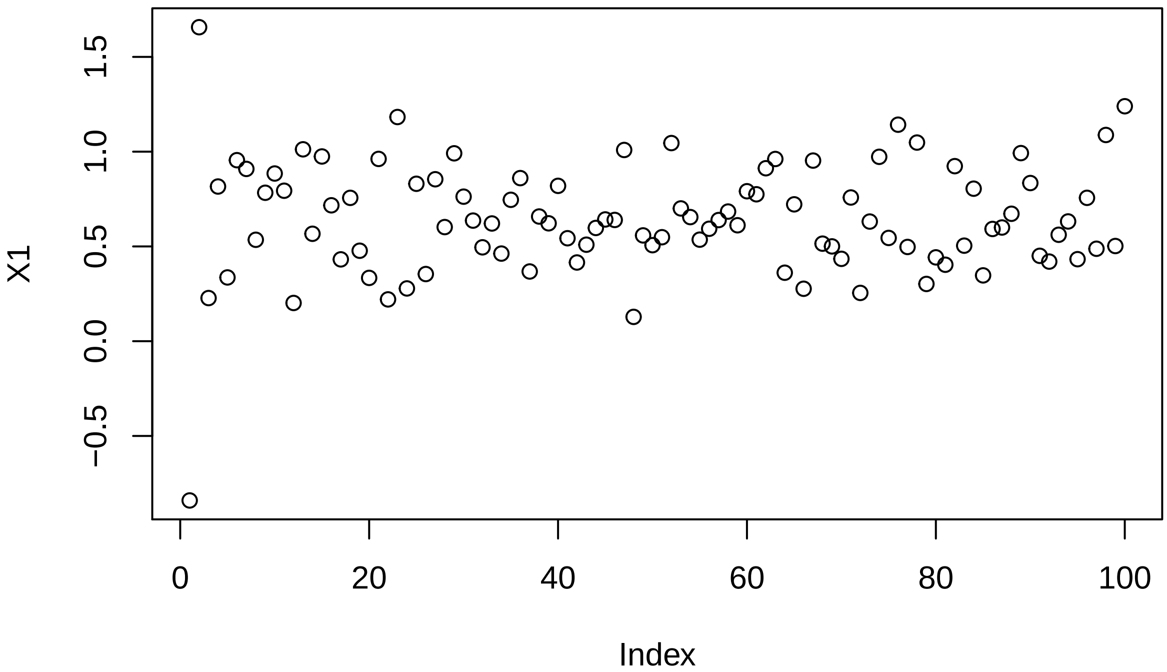
Problem Xtra #14

a)

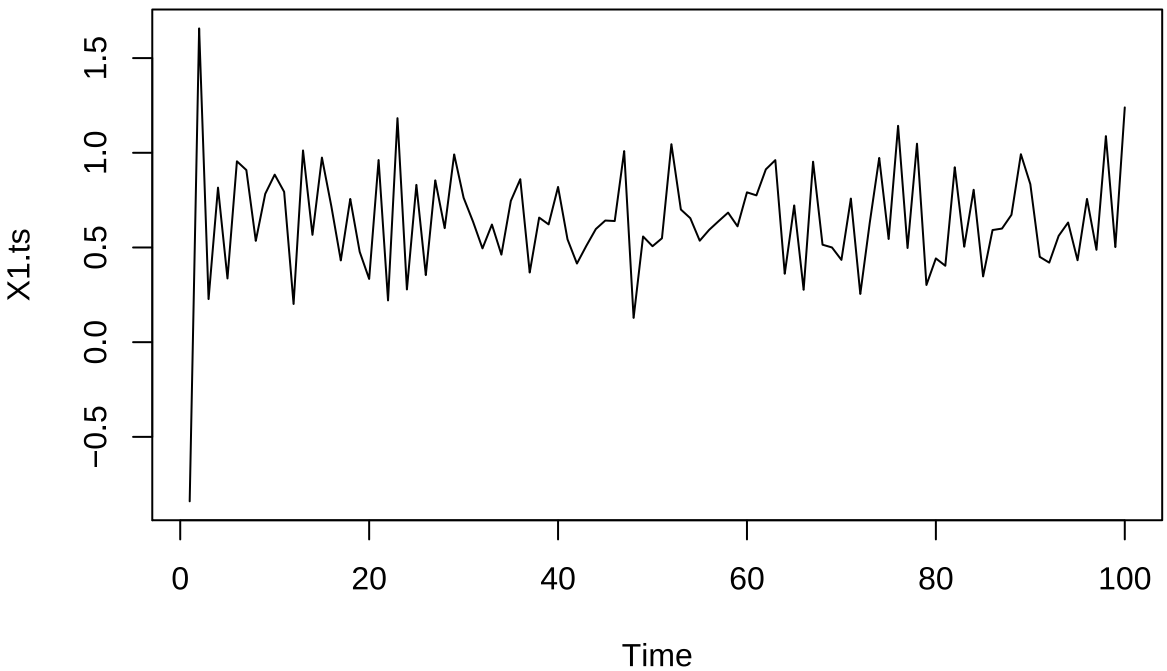
```r
set.seed(134)

# Generate X
X1=rep(NA,100)
X1[1]=rnorm(1)
for (i in 2:100)
{
  X1[i]=1-0.5*X1[i-1]+rnorm(1,0,0.2)
}

# Plot X as a vector
plot(X1)
```




```
# Plot as timeseries object
X1.ts=as.ts(X1)
plot(X1.ts)
```



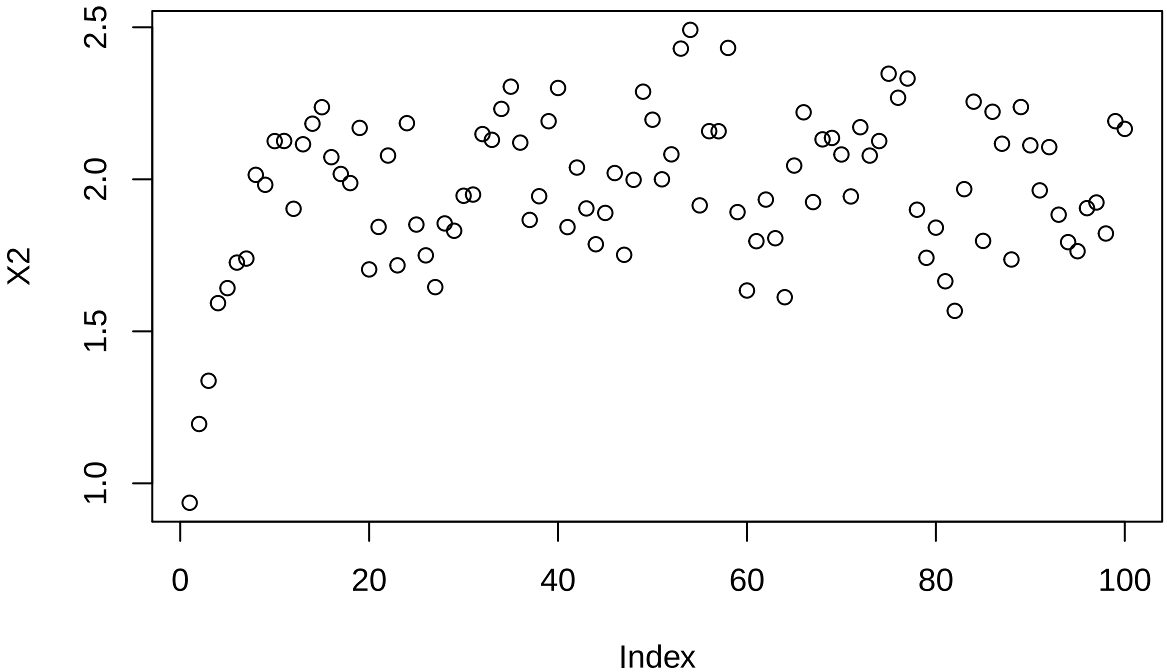
From the first plot we can see that the data points spread equally over time. From the second plot, we can see that X has fluctuation over time.

b)

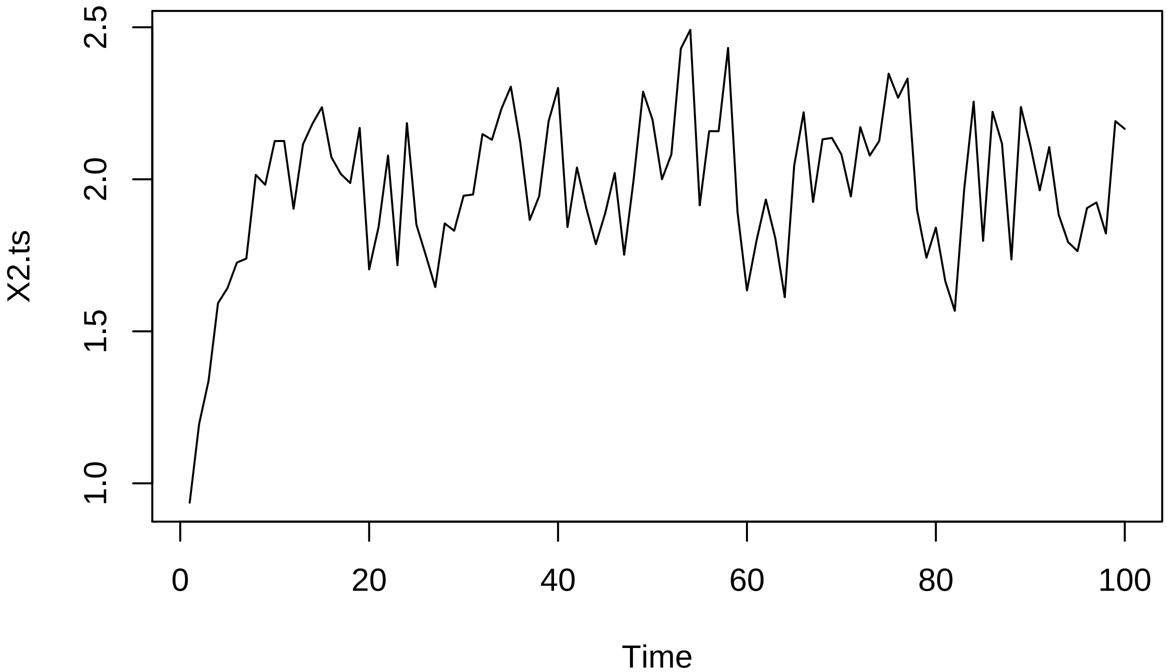
```
# Generate X
X2=rep(NA,100)
```

```r
X2[1]=rnorm(1)
for (i in 2:100)
{
  X2[i]=1+0.5*X2[i-1]+rnorm(1,0,0.2)
}

# Plot X as a vector
plot(X2)
```




```r
# Plot as timeseries object
X2.ts=as.ts(X2)
plot(X2.ts)
```

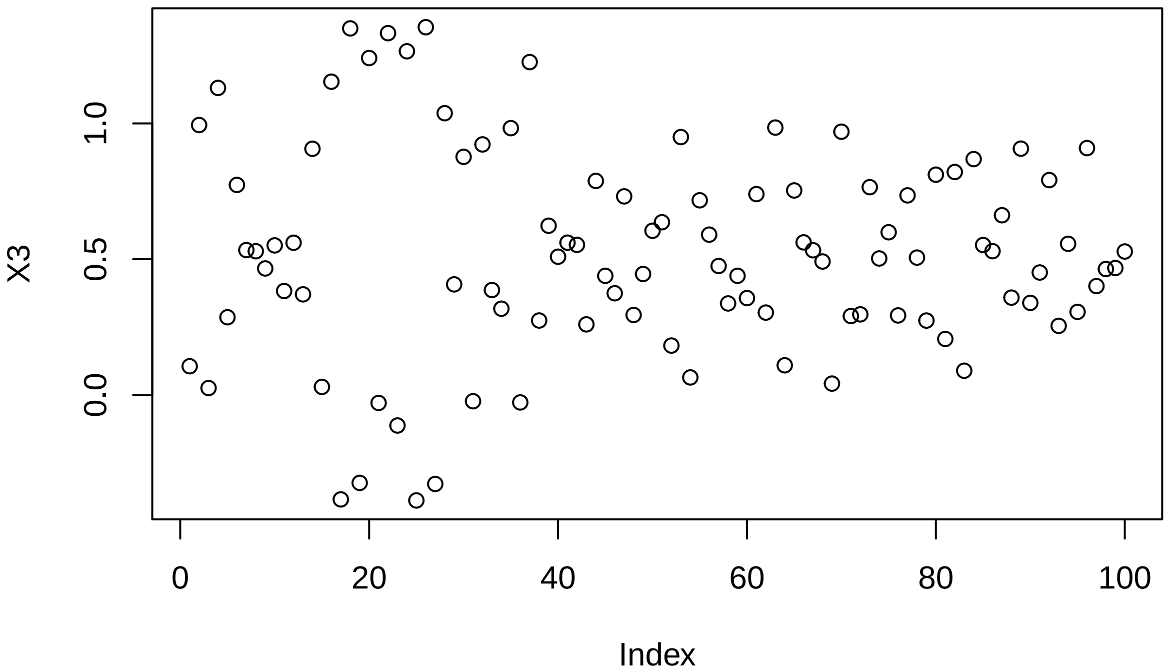


From the first plot we can see that the data points spread much wider than a) over time. From the second plot, we can see that X has bigger fluctuation over time.

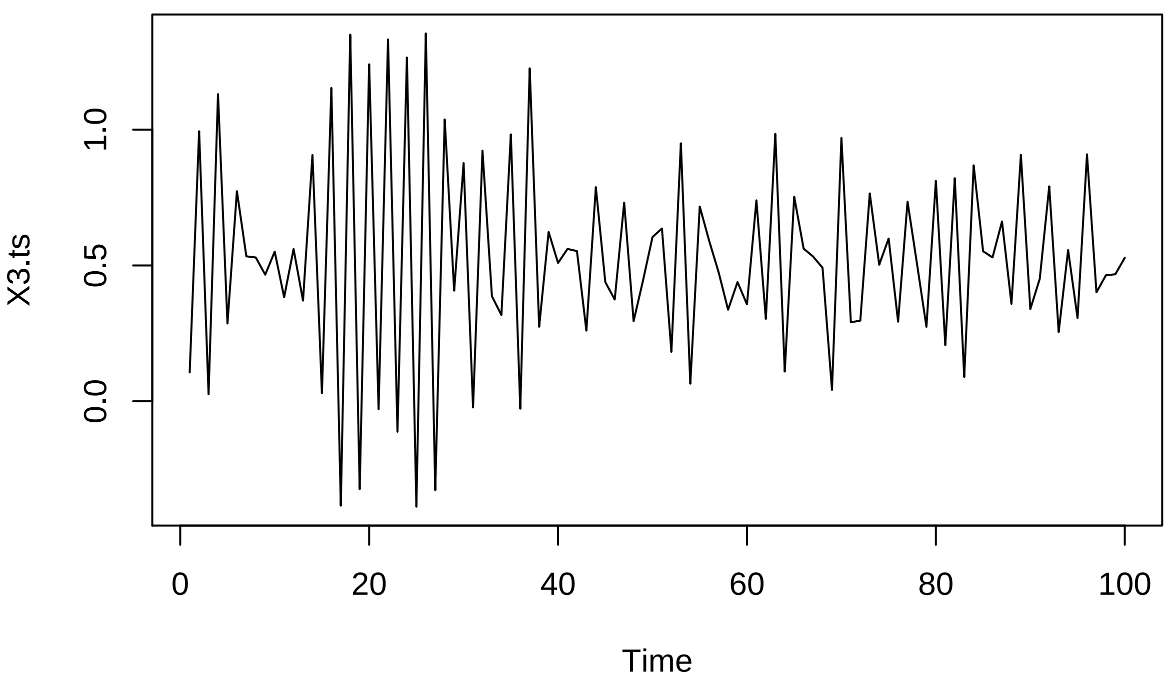
c)

```
# Generate X
X3=rep(NA,100)
X3[1]=rnorm(1)
for (i in 2:100)
{
  X3[i]=1-0.9*X3[i-1]+rnorm(1,0,0.2)
}

# Plot X as a vector
plot(X3)
```




```r
# Plot as timeseries object
X3.ts=as.ts(X3)
plot(X3.ts)
```



From the first plot we can see that the data points spread much wider than b) over time. From the second plot, we can see that X has bigger fluctuation over time and for a given period of time, the number of fluctuations that happend during that time increases.

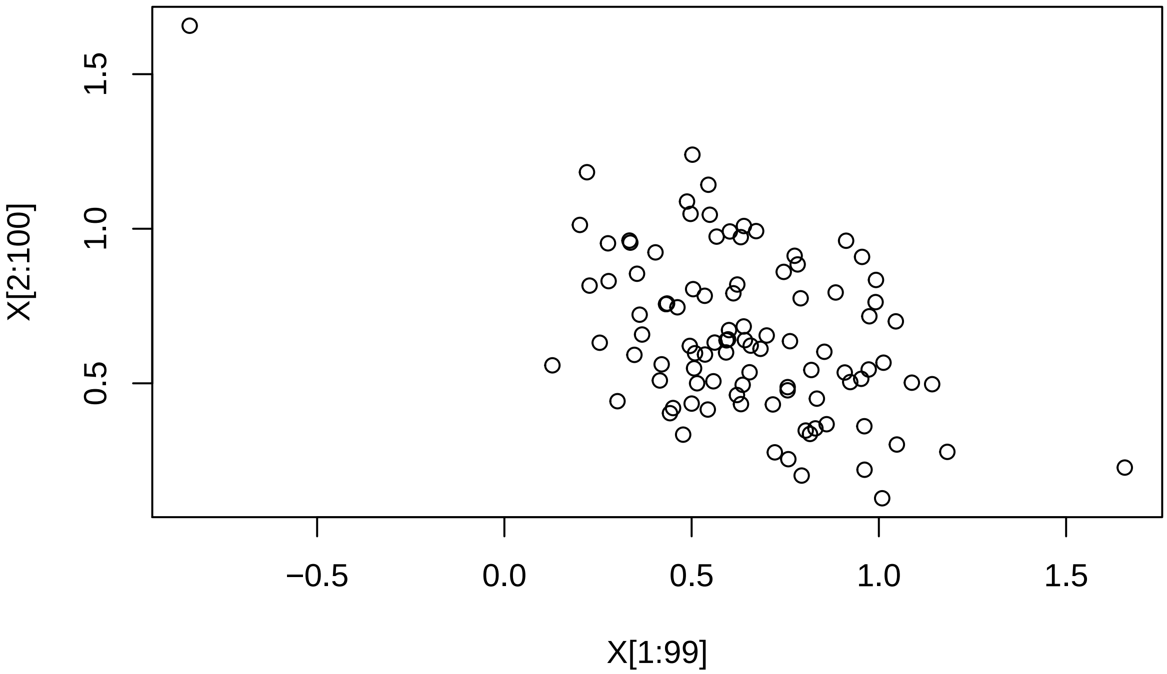
Problem Xtra #15

a)

```
set.seed(134)

# Generate X
X=rep(NA,100)
X[1]=rnorm(1)
for (i in 2:100)
{
  X[i]=1-0.5*X[i-1]+rnorm(1,0,0.2)
}

# plot
plot(X[1:99], X[2:100])
```



It seems that there is a negative relationship between X_{t-1} and X_t.

b)

```
# Data Frame
time.data=data.frame(Xt_1=X[1:99],X=X[2:100])

# Fit a model
time.model=lm(X~Xt_1,data=time.data)
summary(time.model)
```

```
##
## Call:
## lm(formula = X ~ Xt_1, data = time.data)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -0.39977 -0.18074 -0.00513  0.15606  0.51720
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.95359    0.05406  17.640  < 2e-16 ***
## Xt_1        -0.46045    0.07681  -5.995 3.49e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2261 on 97 degrees of freedom
## Multiple R-squared:  0.2703, Adjusted R-squared:  0.2628
## F-statistic: 35.93 on 1 and 97 DF,  p-value: 3.486e-08
```

$\hat{\beta}_i$ is -0.46045, which is close to the real $\beta_i = -0.5$. The residual standard error is 0.2261, which is close to the real standard error 0.2.

The model we bulid is $X_t = 0.95359 - 0.46045 * X_{t-1}$. There is a negative relationship between X_t and X_{t-1}, so when the previous observation changes one unit, X_t will decrease about 0.46 units.