Fin Time Series Homework 2

Zhenzhen Zhang(198000379)

Question 1 a) ¶

```
In [1]: library(MASS)
        data(Boston)
```

```
In [61]: dim(Boston)
```

506 14

```
In [2]: 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
```

```
In [3]: lmod<-lm(medv~crim+zn+indus+chas+nox+rm+age+dis+rad+tax+ptratio+black+lstat,da
        ta=Boston)
```

In [4]: `summary(lmod)`

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

Residuals:
    Min      1Q  Median      3Q     Max
-15.595  -2.730  -0.518   1.777  26.199

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.646e+01  5.103e+00   7.144 3.28e-12 ***
crim        -1.080e-01  3.286e-02  -3.287 0.001087 **
zn           4.642e-02  1.373e-02   3.382 0.000778 ***
indus        2.056e-02  6.150e-02   0.334 0.738288
chas         2.687e+00  8.616e-01   3.118 0.001925 **
nox         -1.777e+01  3.820e+00  -4.651 4.25e-06 ***
rm           3.810e+00  4.179e-01   9.116  < 2e-16 ***
age          6.922e-04  1.321e-02   0.052 0.958229
dis         -1.476e+00  1.995e-01  -7.398 6.01e-13 ***
rad          3.060e-01  6.635e-02   4.613 5.07e-06 ***
tax         -1.233e-02  3.760e-03  -3.280 0.001112 **
ptratio     -9.527e-01  1.308e-01  -7.283 1.31e-12 ***
black        9.312e-03  2.686e-03   3.467 0.000573 ***
lstat       -5.248e-01  5.072e-02 -10.347  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.745 on 492 degrees of freedom
Multiple R-squared:  0.7406,    Adjusted R-squared:  0.7338
F-statistic: 108.1 on 13 and 492 DF,  p-value: < 2.2e-16
```

From the summary of the mulitple regression, we can see that zn,nox,rm,dis,rad,ptratio,black,lstat are statistically significant given a 99.9% confidence interval. Similarly, crim, chas, tax are significant given a 99% confidence interval. For indus and age, they are not statistically significant.

In [21]:
```
step1<-step(lmod,direction="backward")
```

```
Start:   AIC=1589.64
medv ~ crim + zn + indus + chas + nox + rm + age + dis + rad +
    tax + ptratio + black + lstat

          Df Sum of Sq    RSS    AIC
- age      1       0.06  11079  1587.7
- indus    1       2.52  11081  1587.8
<none>                   11079  1589.6
- chas     1     218.97  11298  1597.5
- tax      1     242.26  11321  1598.6
- crim     1     243.22  11322  1598.6
- zn       1     257.49  11336  1599.3
- black    1     270.63  11349  1599.8
- rad      1     479.15  11558  1609.1
- nox      1     487.16  11566  1609.4
- ptratio  1    1194.23  12273  1639.4
- dis      1    1232.41  12311  1641.0
- rm       1    1871.32  12950  1666.6
- lstat    1    2410.84  13490  1687.3

Step:   AIC=1587.65
medv ~ crim + zn + indus + chas + nox + rm + dis + rad + tax +
    ptratio + black + lstat

          Df Sum of Sq    RSS    AIC
- indus    1       2.52  11081  1585.8
<none>                   11079  1587.7
- chas     1     219.91  11299  1595.6
- tax      1     242.24  11321  1596.6
- crim     1     243.20  11322  1596.6
- zn       1     260.32  11339  1597.4
- black    1     272.26  11351  1597.9
- rad      1     481.09  11560  1607.2
- nox      1     520.87  11600  1608.9
- ptratio  1    1200.23  12279  1637.7
- dis      1    1352.26  12431  1643.9
- rm       1    1959.55  13038  1668.0
- lstat    1    2718.88  13798  1696.7

Step:   AIC=1585.76
medv ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +
    black + lstat

          Df Sum of Sq    RSS    AIC
<none>                   11081  1585.8
- chas     1     227.21  11309  1594.0
- crim     1     245.37  11327  1594.8
- zn       1     257.82  11339  1595.4
- black    1     270.82  11352  1596.0
- tax      1     273.62  11355  1596.1
- rad      1     500.92  11582  1606.1
- nox      1     541.91  11623  1607.9
- ptratio  1    1206.45  12288  1636.0
- dis      1    1448.94  12530  1645.9
- rm       1    1963.66  13045  1666.3
- lstat    1    2723.48  13805  1695.0
```

In [22]: `summary(step1)`

```
Call:
lm(formula = medv ~ crim + zn + chas + nox + rm + dis + rad +
    tax + ptratio + black + lstat, data = Boston)

Residuals:
     Min      1Q  Median      3Q     Max
-15.5984 -2.7386 -0.5046  1.7273 26.2373

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  36.341145   5.067492   7.171 2.73e-12 ***
crim         -0.108413   0.032779  -3.307 0.001010 **
zn            0.045845   0.013523   3.390 0.000754 ***
chas          2.718716   0.854240   3.183 0.001551 **
nox         -17.376023   3.535243  -4.915 1.21e-06 ***
rm            3.801579   0.406316   9.356  < 2e-16 ***
dis          -1.492711   0.185731  -8.037 6.84e-15 ***
rad           0.299608   0.063402   4.726 3.00e-06 ***
tax          -0.011778   0.003372  -3.493 0.000521 ***
ptratio      -0.946525   0.129066  -7.334 9.24e-13 ***
black         0.009291   0.002674   3.475 0.000557 ***
lstat        -0.522553   0.047424 -11.019  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.736 on 494 degrees of freedom
Multiple R-squared:  0.7406,    Adjusted R-squared:  0.7348
F-statistic: 128.2 on 11 and 494 DF,  p-value: < 2.2e-16
```

In [23]:
```
step2<-step(lmod,direction="backward", k=log(nrow(Boston)))
```

```
Start:  AIC=1648.81
medv ~ crim + zn + indus + chas + nox + rm + age + dis + rad +
    tax + ptratio + black + lstat

          Df Sum of Sq   RSS    AIC
- age      1       0.06 11079 1642.6
- indus    1       2.52 11081 1642.7
<none>                  11079 1648.8
- chas     1     218.97 11298 1652.5
- tax      1     242.26 11321 1653.5
- crim     1     243.22 11322 1653.6
- zn       1     257.49 11336 1654.2
- black    1     270.63 11349 1654.8
- rad      1     479.15 11558 1664.0
- nox      1     487.16 11566 1664.4
- ptratio  1    1194.23 12273 1694.4
- dis      1    1232.41 12311 1696.0
- rm       1    1871.32 12950 1721.6
- lstat    1    2410.84 13490 1742.2

Step:  AIC=1642.59
medv ~ crim + zn + indus + chas + nox + rm + dis + rad + tax +
    ptratio + black + lstat

          Df Sum of Sq   RSS    AIC
- indus    1       2.52 11081 1636.5
<none>                  11079 1642.6
- chas     1     219.91 11299 1646.3
- tax      1     242.24 11321 1647.3
- crim     1     243.20 11322 1647.3
- zn       1     260.32 11339 1648.1
- black    1     272.26 11351 1648.7
- rad      1     481.09 11560 1657.9
- nox      1     520.87 11600 1659.6
- ptratio  1    1200.23 12279 1688.4
- dis      1    1352.26 12431 1694.6
- rm       1    1959.55 13038 1718.8
- lstat    1    2718.88 13798 1747.4

Step:  AIC=1636.48
medv ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +
    black + lstat

          Df Sum of Sq   RSS    AIC
<none>                  11081 1636.5
- chas     1     227.21 11309 1640.5
- crim     1     245.37 11327 1641.3
- zn       1     257.82 11339 1641.9
- black    1     270.82 11352 1642.5
- tax      1     273.62 11355 1642.6
- rad      1     500.92 11582 1652.6
- nox      1     541.91 11623 1654.4
- ptratio  1    1206.45 12288 1682.5
- dis      1    1448.94 12530 1692.4
- rm       1    1963.66 13045 1712.8
- lstat    1    2723.48 13805 1741.5
```

In [26]: `summary(step2)`

```
Call:
lm(formula = medv ~ crim + zn + chas + nox + rm + dis + rad +
    tax + ptratio + black + lstat, data = Boston)

Residuals:
     Min       1Q   Median       3Q      Max
-15.5984  -2.7386  -0.5046   1.7273  26.2373

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  36.341145   5.067492   7.171 2.73e-12 ***
crim         -0.108413   0.032779  -3.307 0.001010 **
zn            0.045845   0.013523   3.390 0.000754 ***
chas          2.718716   0.854240   3.183 0.001551 **
nox         -17.376023   3.535243  -4.915 1.21e-06 ***
rm            3.801579   0.406316   9.356  < 2e-16 ***
dis          -1.492711   0.185731  -8.037 6.84e-15 ***
rad           0.299608   0.063402   4.726 3.00e-06 ***
tax          -0.011778   0.003372  -3.493 0.000521 ***
ptratio      -0.946525   0.129066  -7.334 9.24e-13 ***
black         0.009291   0.002674   3.475 0.000557 ***
lstat        -0.522553   0.047424 -11.019  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.736 on 494 degrees of freedom
Multiple R-squared:  0.7406,    Adjusted R-squared:  0.7348
F-statistic: 128.2 on 11 and 494 DF,  p-value: < 2.2e-16
```
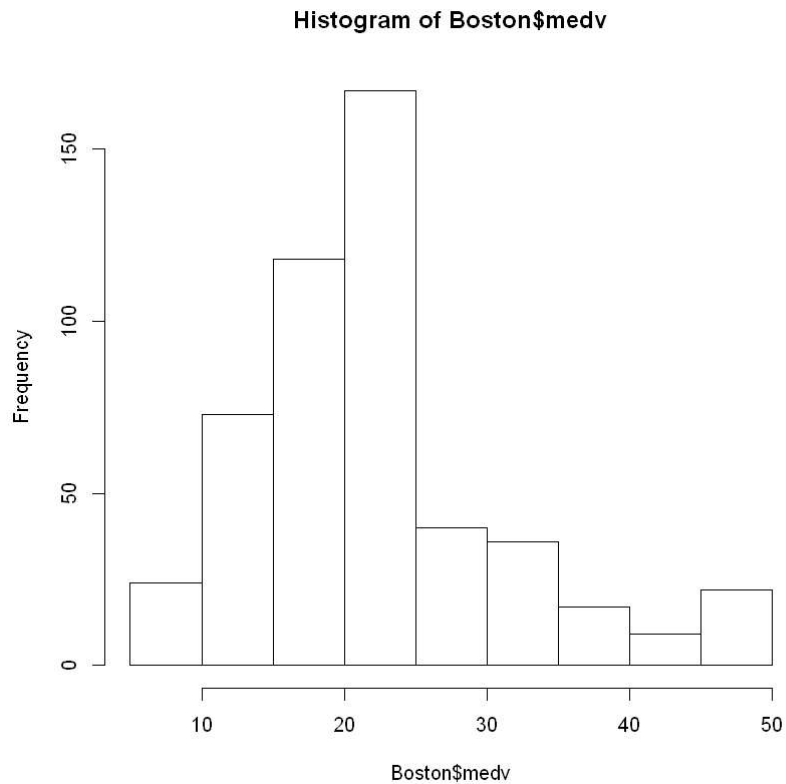
(2) From the AIC and BIC, we can see that both AIC and BIC set indus and age aside the model and the remaining variables are all significant.
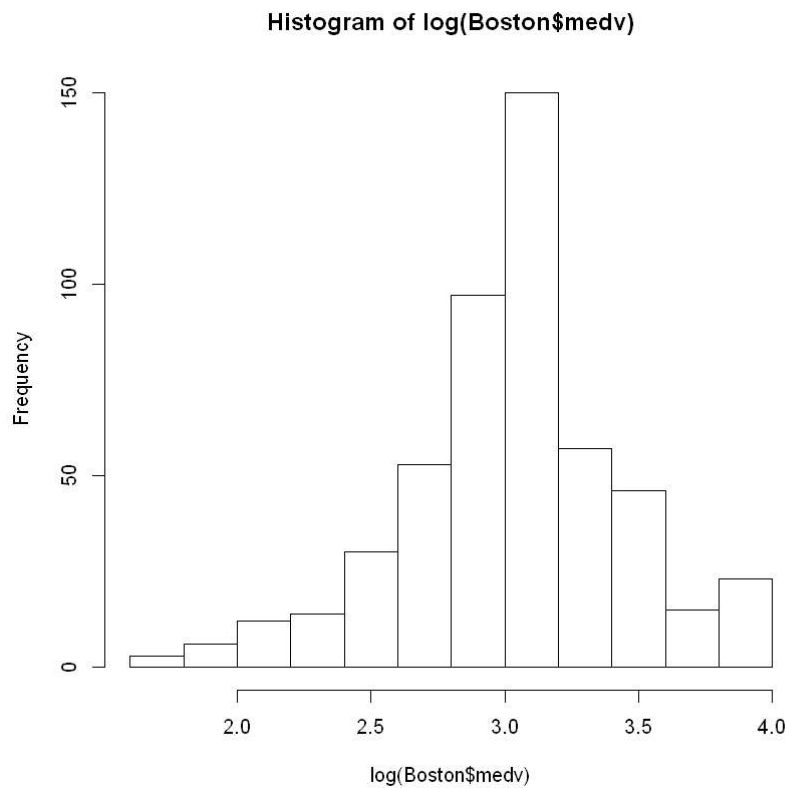
Question 1 b)

In [14]: `hist(Boston$medv)`

Histogram of Boston$medv

From the histogram we can know that medv is right-skewed because the distribution's peak is off center toward the limit and a tail stretches away from it. In this case, we need to modify the outcome variable. For example, log(medv):

In [58]: `hist(log(Boston$medv))`

Histogram of log(Boston$medv)

In [18]: `lmod2<-lm(log(medv)~crim+zn+indus+chas+nox+rm+age+dis+rad+tax+ptratio+black+lstat,data=Boston)`

In [20]: `summary(lmod2)`

```
Call:
lm(formula = log(medv) ~ crim + zn + indus + chas + nox + rm +
    age + dis + rad + tax + ptratio + black + lstat, data = Boston)

Residuals:
     Min       1Q   Median       3Q      Max
-0.73361 -0.09747 -0.01657  0.09629  0.86435

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.1020423  0.2042726  20.081  < 2e-16 ***
crim        -0.0102715  0.0013155  -7.808 3.52e-14 ***
zn           0.0011725  0.0005495   2.134 0.033349 *
indus        0.0024668  0.0024614   1.002 0.316755
chas         0.1008876  0.0344859   2.925 0.003598 **
nox         -0.7783993  0.1528902  -5.091 5.07e-07 ***
rm           0.0908331  0.0167280   5.430 8.87e-08 ***
age          0.0002106  0.0005287   0.398 0.690567
dis         -0.0490873  0.0079834  -6.149 1.62e-09 ***
rad          0.0142673  0.0026556   5.373 1.20e-07 ***
tax         -0.0006258  0.0001505  -4.157 3.80e-05 ***
ptratio     -0.0382715  0.0052365  -7.309 1.10e-12 ***
black        0.0004136  0.0001075   3.847 0.000135 ***
lstat       -0.0290355  0.0020299 -14.304  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1899 on 492 degrees of freedom
Multiple R-squared:  0.7896,    Adjusted R-squared:  0.7841
F-statistic: 142.1 on 13 and 492 DF,  p-value: < 2.2e-16
```

```
In [27]: step3<-step(lmod2,direction="backward")
```

```
Start:  AIC=-1667.19
log(medv) ~ crim + zn + indus + chas + nox + rm + age + dis +
    rad + tax + ptratio + black + lstat

          Df Sum of Sq     RSS      AIC
- age      1     0.0057  17.755  -1669.0
- indus    1     0.0362  17.786  -1668.2
<none>                   17.749  -1667.2
- zn       1     0.1643  17.914  -1664.5
- chas     1     0.3088  18.058  -1660.5
- black    1     0.5339  18.283  -1654.2
- tax      1     0.6235  18.373  -1651.7
- nox      1     0.9351  18.684  -1643.2
- rad      1     1.0413  18.791  -1640.3
- rm       1     1.0637  18.813  -1639.7
- dis      1     1.3639  19.113  -1631.7
- ptratio  1     1.9270  19.676  -1617.0
- crim     1     2.1995  19.949  -1610.1
- lstat    1     7.3809  25.130  -1493.2

Step:  AIC=-1669.03
log(medv) ~ crim + zn + indus + chas + nox + rm + dis + rad +
    tax + ptratio + black + lstat

          Df Sum of Sq     RSS      AIC
- indus    1     0.0363  17.791  -1670.0
<none>                   17.755  -1669.0
- zn       1     0.1593  17.914  -1666.5
- chas     1     0.3138  18.069  -1662.2
- black    1     0.5431  18.298  -1655.8
- tax      1     0.6205  18.376  -1653.7
- nox      1     0.9645  18.720  -1644.3
- rad      1     1.0356  18.791  -1642.3
- rm       1     1.1452  18.900  -1639.4
- dis      1     1.5471  19.302  -1628.8
- ptratio  1     1.9224  19.677  -1619.0
- crim     1     2.1988  19.954  -1612.0
- lstat    1     8.1949  25.950  -1479.0

Step:  AIC=-1670
log(medv) ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +
    black + lstat

          Df Sum of Sq     RSS      AIC
<none>                   17.791  -1670.0
- zn       1     0.1451  17.936  -1667.9
- chas     1     0.3399  18.131  -1662.4
- black    1     0.5344  18.326  -1657.0
- tax      1     0.6139  18.405  -1654.8
- nox      1     0.9350  18.726  -1646.1
- rad      1     1.0088  18.800  -1644.1
- rm       1     1.1171  18.909  -1641.2
- dis      1     1.7385  19.530  -1624.8
- ptratio  1     1.8862  19.678  -1621.0
- crim     1     2.2229  20.014  -1612.4
- lstat    1     8.1604  25.952  -1481.0
```

In [28]: `summary(step3)`

```
Call:
lm(formula = log(medv) ~ crim + zn + chas + nox + rm + dis +
    rad + tax + ptratio + black + lstat, data = Boston)

Residuals:
     Min      1Q   Median      3Q      Max
-0.73400 -0.09460 -0.01771  0.09782  0.86290

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.0836823  0.2030491  20.112  < 2e-16 ***
crim        -0.0103187  0.0013134  -7.856 2.49e-14 ***
zn           0.0010874  0.0005418   2.007 0.045308 *
chas         0.1051484  0.0342285   3.072 0.002244 **
nox         -0.7217440  0.1416535  -5.095 4.97e-07 ***
rm           0.0906728  0.0162807   5.569 4.20e-08 ***
dis         -0.0517059  0.0074420  -6.948 1.18e-11 ***
rad          0.0134457  0.0025405   5.293 1.82e-07 ***
tax         -0.0005579  0.0001351  -4.129 4.28e-05 ***
ptratio     -0.0374259  0.0051715  -7.237 1.77e-12 ***
black        0.0004127  0.0001071   3.852 0.000133 ***
lstat       -0.0286039  0.0019002 -15.053  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1898 on 494 degrees of freedom
Multiple R-squared:  0.7891,    Adjusted R-squared:  0.7844
F-statistic: 168.1 on 11 and 494 DF,  p-value: < 2.2e-16
```

In [29]:
```r
step4<-step(lmod2,direction="backward", k=log(nrow(Boston)))
```

```
Start:  AIC=-1608.02
log(medv) ~ crim + zn + indus + chas + nox + rm + age + dis +
    rad + tax + ptratio + black + lstat

          Df Sum of Sq    RSS      AIC
- age      1     0.0057 17.755 -1614.1
- indus    1     0.0362 17.786 -1613.2
- zn       1     0.1643 17.914 -1609.6
<none>                  17.749 -1608.0
- chas     1     0.3088 18.058 -1605.5
- black    1     0.5339 18.283 -1599.2
- tax      1     0.6235 18.373 -1596.8
- nox      1     0.9351 18.684 -1588.3
- rad      1     1.0413 18.791 -1585.4
- rm       1     1.0637 18.813 -1584.8
- dis      1     1.3639 19.113 -1576.8
- ptratio  1     1.9270 19.676 -1562.1
- crim     1     2.1995 19.949 -1555.1
- lstat    1     7.3809 25.130 -1438.3

Step:  AIC=-1614.09
log(medv) ~ crim + zn + indus + chas + nox + rm + dis + rad +
    tax + ptratio + black + lstat

          Df Sum of Sq    RSS      AIC
- indus    1     0.0363 17.791 -1619.3
- zn       1     0.1593 17.914 -1615.8
<none>                  17.755 -1614.1
- chas     1     0.3138 18.069 -1611.5
- black    1     0.5431 18.298 -1605.1
- tax      1     0.6205 18.376 -1602.9
- nox      1     0.9645 18.720 -1593.5
- rad      1     1.0356 18.791 -1591.6
- rm       1     1.1452 18.900 -1588.7
- dis      1     1.5471 19.302 -1578.0
- ptratio  1     1.9224 19.677 -1568.3
- crim     1     2.1988 19.954 -1561.2
- lstat    1     8.1949 25.950 -1428.3

Step:  AIC=-1619.28
log(medv) ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +
    black + lstat

          Df Sum of Sq    RSS      AIC
- zn       1     0.1451 17.936 -1621.4
<none>                  17.791 -1619.3
- chas     1     0.3399 18.131 -1615.9
- black    1     0.5344 18.326 -1610.5
- tax      1     0.6139 18.405 -1608.3
- nox      1     0.9350 18.726 -1599.6
- rad      1     1.0088 18.800 -1597.6
- rm       1     1.1171 18.909 -1594.7
- dis      1     1.7385 19.530 -1578.3
- ptratio  1     1.8862 19.678 -1574.5
- crim     1     2.2229 20.014 -1565.9
- lstat    1     8.1604 25.952 -1434.5
```

```
Step:  AIC=-1621.4
log(medv) ~ crim + chas + nox + rm + dis + rad + tax + ptratio +
    black + lstat

          Df Sum of Sq    RSS      AIC
<none>                  17.936 -1621.4
- chas     1    0.3388 18.275 -1618.2
- tax      1    0.5229 18.459 -1613.1
- black    1    0.5386 18.475 -1612.7
- rad      1    0.9601 18.897 -1601.2
- nox      1    1.0250 18.961 -1599.5
- rm       1    1.2650 19.201 -1593.1
- dis      1    1.6967 19.633 -1581.9
- crim     1    2.1377 20.074 -1570.7
- ptratio  1    2.5632 20.500 -1560.0
- lstat    1    8.1516 26.088 -1438.1
```

In [30]:   `summary(step4)`

```
Call:
lm(formula = log(medv) ~ crim + chas + nox + rm + dis + rad +
    tax + ptratio + black + lstat, data = Boston)

Residuals:
     Min       1Q   Median       3Q      Max
-0.73091 -0.10531 -0.01614  0.09871  0.87074

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.1000969  0.2035038  20.148  < 2e-16 ***
crim        -0.0100763  0.0013119  -7.681 8.52e-14 ***
chas         0.1049848  0.0343329   3.058 0.002350 **
nox         -0.7515379  0.1413035  -5.319 1.59e-07 ***
rm           0.0954516  0.0161547   5.909 6.43e-09 ***
dis         -0.0442395  0.0064650  -6.843 2.30e-11 ***
rad          0.0130841  0.0025418   5.148 3.81e-07 ***
tax         -0.0005050  0.0001329  -3.799 0.000164 ***
ptratio     -0.0409840  0.0048729  -8.411 4.39e-16 ***
black        0.0004143  0.0001075   3.855 0.000131 ***
lstat       -0.0285881  0.0019060 -14.999  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1904 on 495 degrees of freedom
Multiple R-squared:  0.7874,    Adjusted R-squared:  0.7831
F-statistic: 183.4 on 10 and 495 DF,  p-value: < 2.2e-16
```

From step1 to step4, these all don't use indus and age in models. And there are some obvious differences between the outcomes. To be specific, for the remaining varibles: step1 in AIC: medv ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio + black + lstat step2 in BIC: medv ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio + black + lstat step3 in AIC: log(medv) ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio + black + lstat step4 in BIC: log(medv) ~ crim + chas + nox + rm + dis + rad + tax + ptratio + black + lstat Only step4 removes zn in model.

For adjusted R2: step1 in AIC: 0.7348 step2 in BIC: 0.7348 step3 in AIC: 0.7844 step4 in BIC: 0.7831 From above, we can know that step3 can explain more in outcome so it's better than other three.

Question 2 a)

```
In [81]:  install.packages("fpp")
          library("fpp")
          data(fancy)
```

```
Installing package into 'C:/Users/zzzha/Documents/R/win-library/3.6'
(as 'lib' is unspecified)

Warning message:
"package 'fpp' is in use and will not be installed"
Warning message in data(fancy):
"data set 'fancy' not found"
```

```
In [38]:  summary(fancy)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1665    5884    8772   14316   16889  104661
```

```
In [60]:  fancy
```

A Time Series: 7 × 12

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
1987	1664.81	2397.53	2840.71	3547.29	3752.96	3714.74	4349.61	3566.34	5021.82
1988	2499.81	5198.24	7225.14	4806.03	5900.88	4951.34	6179.12	4752.15	5496.43
1989	4717.02	5702.63	9957.58	5304.78	6492.43	6630.80	7349.62	8176.62	8573.17
1990	5921.10	5814.58	12421.25	6369.77	7609.12	7224.75	8121.22	7979.25	8093.06
1991	4826.64	6470.23	9638.77	8821.17	8722.37	10209.48	11276.55	12552.22	11637.39
1992	7615.03	9849.69	14558.40	11587.33	9332.56	13082.09	16732.78	19888.61	23933.38
1993	10243.24	11266.88	21826.84	17357.33	15997.79	18601.53	26155.15	28586.52	30505.41

In [40]: `plot.ts(fancy)`
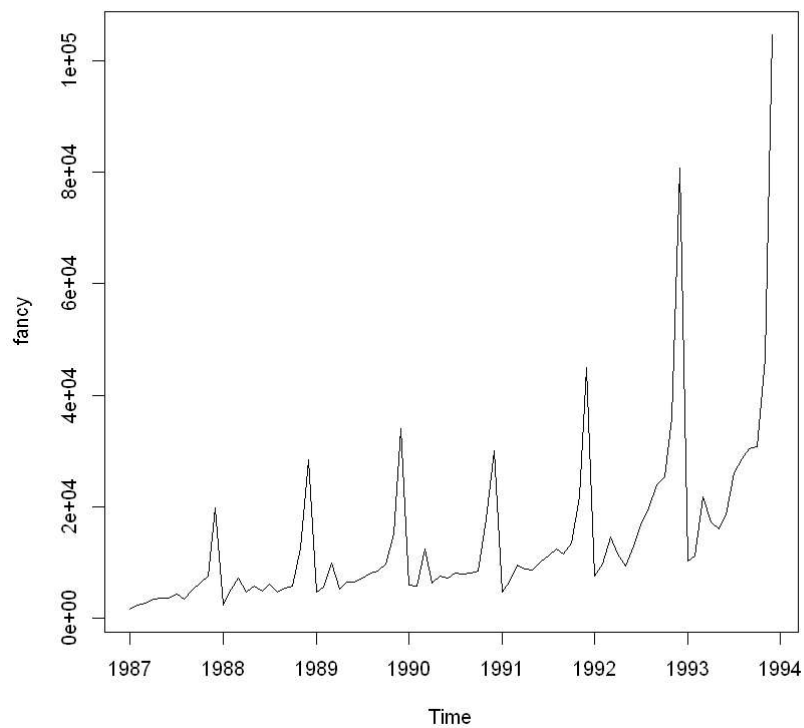
From the summary(fancy) and the plot, we can know that there is a increase in monthly sales every March and December. Also, the increase is gradually growing while 1991 is an exception.

Question 2 b)

In [87]: `y <- log(fancy)`

In [88]: `fit <- tslm(y ~ trend + season,data=y)`

In [89]: `summary(fit)`

```
Call:
tslm(formula = y ~ trend + season, data = y)

Residuals:
     Min       1Q   Median       3Q      Max
-0.41644 -0.12619  0.00608  0.11389  0.38567

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 7.6058604  0.0768740  98.939  < 2e-16 ***
trend       0.0223930  0.0008448  26.508  < 2e-16 ***
season2     0.2510437  0.0993278   2.527 0.013718 *
season3     0.6952066  0.0993386   6.998 1.18e-09 ***
season4     0.3829341  0.0993565   3.854 0.000252 ***
season5     0.4079944  0.0993817   4.105 0.000106 ***
season6     0.4469625  0.0994140   4.496 2.63e-05 ***
season7     0.6082156  0.0994534   6.116 4.69e-08 ***
season8     0.5853524  0.0995001   5.883 1.21e-07 ***
season9     0.6663446  0.0995538   6.693 4.27e-09 ***
season10    0.7440336  0.0996148   7.469 1.61e-10 ***
season11    1.2030164  0.0996828  12.068  < 2e-16 ***
season12    1.9581366  0.0997579  19.629  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1858 on 71 degrees of freedom
Multiple R-squared:  0.9527,    Adjusted R-squared:  0.9447
F-statistic: 119.1 on 12 and 71 DF,  p-value: < 2.2e-16
```

Question 2 c)

In [82]:
```
x1=1:84
x2=rep(c(0,1,0,0,0,0,0,0,0,0,0,0),7)
x3=rep(c(0,0,1,0,0,0,0,0,0,0,0,0),7)
x4=rep(c(0,0,0,1,0,0,0,0,0,0,0,0),7)
x5=rep(c(0,0,0,0,1,0,0,0,0,0,0,0),7)
x6=rep(c(0,0,0,0,0,1,0,0,0,0,0,0),7)
x7=rep(c(0,0,0,0,0,0,1,0,0,0,0,0),7)
x8=rep(c(0,0,0,0,0,0,0,1,0,0,0,0),7)
x9=rep(c(0,0,0,0,0,0,0,0,1,0,0,0),7)
x10=rep(c(0,0,0,0,0,0,0,0,0,1,0,0),7)
x11=rep(c(0,0,0,0,0,0,0,0,0,0,1,0),7)
x12=rep(c(0,0,0,0,0,0,0,0,0,0,0,1),7)
```

In [85]:
```
fit1<-lm(log(fancy)~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12)
```

In [86]: `summary(fit1)`

```
Call:
lm(formula = log(fancy) ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 +
    x8 + x9 + x10 + x11 + x12)

Residuals:
      Min       1Q    Median       3Q      Max
 -0.41644 -0.12619   0.00608  0.11389  0.38567

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 7.6058604  0.0768740  98.939  < 2e-16 ***
x1          0.0223930  0.0008448  26.508  < 2e-16 ***
x2          0.2510437  0.0993278   2.527 0.013718 *
x3          0.6952066  0.0993386   6.998 1.18e-09 ***
x4          0.3829341  0.0993565   3.854 0.000252 ***
x5          0.4079944  0.0993817   4.105 0.000106 ***
x6          0.4469625  0.0994140   4.496 2.63e-05 ***
x7          0.6082156  0.0994534   6.116 4.69e-08 ***
x8          0.5853524  0.0995001   5.883 1.21e-07 ***
x9          0.6663446  0.0995538   6.693 4.27e-09 ***
x10         0.7440336  0.0996148   7.469 1.61e-10 ***
x11         1.2030164  0.0996828  12.068  < 2e-16 ***
x12         1.9581366  0.0997579  19.629  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1858 on 71 degrees of freedom
Multiple R-squared:  0.9527,     Adjusted R-squared:  0.9447
F-statistic: 119.1 on 12 and 71 DF,  p-value: < 2.2e-16
```

From above, we can see that we get same model.