Discussion 6

Xuanpei Ouyang (Esther)

March 21, 2017

(1) Do forward and Backward model selection using AIC. Are the model obtained from forward and backward selection are same?

```
housing_boston = read_csv("~/Desktop/STAT 151A/STAT-151A/lab/lab6/housing_boston.csv")

## Parsed with column specification:
## cols(
##   `0.00632` = col_double(),
##   `18` = col_double(),
##   `2.31` = col_double(),
##   `0` = col_integer(),
##   `0.538` = col_double(),
##   `6.575` = col_double(),
##   `65.2` = col_double(),
##   `4.09` = col_double(),
##   `1` = col_integer(),
##   `296` = col_integer(),
##   `15.3` = col_double(),
##   `396.9` = col_double(),
##   `4.98` = col_double(),
##   `24` = col_double()
## )

colname = c("CRIM","ZN","INDUS","CHAS","NOX","RM","AGE","DIS","RAD","TAX","PTRATIO","B","LSTAT","MEDV")
names(housing_boston) = colname
summary(housing_boston)

##      CRIM               ZN             INDUS            CHAS
##  Min.   : 0.00906   Min.   :  0.00   Min.   : 0.46   Min.   :0.00000
##  1st Qu.: 0.08221   1st Qu.:  0.00   1st Qu.: 5.19   1st Qu.:0.00000
##  Median : 0.25915   Median :  0.00   Median : 9.69   Median :0.00000
##  Mean   : 3.62066   Mean   : 11.35   Mean   :11.15   Mean   :0.06931
##  3rd Qu.: 3.67820   3rd Qu.: 12.50   3rd Qu.:18.10   3rd Qu.:0.00000
##  Max.   :88.97600   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.:5.885   1st Qu.: 45.00   1st Qu.: 2.100
##  Median :0.5380   Median :6.208   Median : 77.70   Median : 3.199
##  Mean   :0.5547   Mean   :6.284   Mean   : 68.58   Mean   : 3.794
##  3rd Qu.:0.6240   3rd Qu.:6.625   3rd Qu.: 94.10   3rd Qu.: 5.212
##  Max.   :0.8710   Max.   :8.780   Max.   :100.00   Max.   :12.127
##      RAD              TAX           PTRATIO            B
##  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.:375.33
##  Median : 5.000   Median :330.0   Median :19.10   Median :391.43
##  Mean   : 9.566   Mean   :408.5   Mean   :18.46   Mean   :356.59
##  3rd Qu.:24.000   3rd Qu.:666.0   3rd Qu.:20.20   3rd Qu.:396.21
##  Max.   :24.000   Max.   :711.0   Max.   :22.00   Max.   :396.90
```

```
##       LSTAT            MEDV
##  Min.   : 1.73   Min.   : 5.00
##  1st Qu.: 7.01   1st Qu.:17.00
##  Median :11.38   Median :21.20
##  Mean   :12.67   Mean   :22.53
##  3rd Qu.:16.96   3rd Qu.:25.00
##  Max.   :37.97   Max.   :50.00
```

```
lm_object = MEDV~CRIM+ZN+INDUS+CHAS+NOX+RM+AGE+DIS+RAD+TAX+PTRATIO+B+LSTAT
housing_boston_model = lm(lm_object, housing_boston)

AIC_forward = step(lm(MEDV ~ 1, data = housing_boston), lm_object, direction = "forward")
```

```
## Start:  AIC=2243.05
## MEDV ~ 1
##
##           Df Sum of Sq   RSS    AIC
## + LSTAT    1   23275.8 19438 1847.5
## + RM       1   20653.6 22060 1911.4
## + PTRATIO  1   11040.8 31673 2094.0
## + INDUS    1   10011.2 32703 2110.2
## + TAX      1    9377.1 33337 2119.9
## + NOX      1    7798.8 34915 2143.2
## + CRIM     1    6438.6 36276 2162.5
## + RAD      1    6222.9 36491 2165.5
## + AGE      1    6068.7 36645 2167.7
## + ZN       1    5547.9 37166 2174.8
## + B        1    4747.7 37966 2185.6
## + DIS      1    2667.4 40047 2212.5
## + CHAS     1    1313.6 41401 2229.3
## <none>                 42714 2243.1
##
## Step:  AIC=1847.47
## MEDV ~ LSTAT
##
##           Df Sum of Sq   RSS    AIC
## + RM       1    4023.5 15415 1732.3
## + PTRATIO  1    2707.8 16730 1773.7
## + CHAS     1     781.7 18657 1828.8
## + DIS      1     779.3 18659 1828.8
## + AGE      1     310.5 19128 1841.3
## + TAX      1     275.2 19163 1842.3
## + B        1     198.7 19240 1844.3
## + ZN       1     159.2 19279 1845.3
## + CRIM     1     146.4 19292 1845.7
## + INDUS    1     103.1 19335 1846.8
## <none>                 19438 1847.5
## + RAD      1      26.5 19412 1848.8
## + NOX      1       5.5 19433 1849.3
##
## Step:  AIC=1732.35
## MEDV ~ LSTAT + RM
##
##           Df Sum of Sq   RSS    AIC
## + PTRATIO  1    1738.57 13676 1673.9
```

```
## + CHAS    1     545.44 14869 1716.2
## + B       1     512.40 14902 1717.3
## + TAX     1     425.87 14989 1720.2
## + DIS     1     355.52 15059 1722.6
## + CRIM    1     310.60 15104 1724.1
## + RAD     1     183.44 15231 1728.3
## + INDUS   1      64.00 15351 1732.2
## <none>               15415 1732.3
## + ZN      1      56.06 15359 1732.5
## + AGE     1      21.70 15393 1733.6
## + NOX     1      13.84 15401 1733.9
##
## Step:  AIC=1673.92
## MEDV ~ LSTAT + RM + PTRATIO
##
##           Df Sum of Sq   RSS    AIC
## + DIS     1     508.25 13168 1656.8
## + B       1     388.72 13288 1661.4
## + CHAS    1     372.75 13304 1662.0
## + CRIM    1     120.37 13556 1671.5
## + AGE     1      70.86 13605 1673.3
## <none>               13676 1673.9
## + TAX     1      42.82 13633 1674.3
## + NOX     1      22.92 13653 1675.1
## + ZN      1      16.23 13660 1675.3
## + RAD     1       6.03 13670 1675.7
## + INDUS   1       0.53 13676 1675.9
##
## Step:  AIC=1656.8
## MEDV ~ LSTAT + RM + PTRATIO + DIS
##
##           Df Sum of Sq   RSS    AIC
## + NOX     1     754.60 12413 1629.0
## + B       1     502.53 12665 1639.2
## + CHAS    1     261.62 12906 1648.7
## + INDUS   1     256.19 12912 1648.9
## + TAX     1     238.61 12929 1649.6
## + CRIM    1     231.41 12937 1649.8
## + ZN      1     143.45 13024 1653.3
## + AGE     1      58.18 13110 1656.6
## <none>               13168 1656.8
## + RAD     1      23.17 13145 1657.9
##
## Step:  AIC=1629
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX
##
##           Df Sum of Sq   RSS    AIC
## + CHAS    1     321.93 12091 1617.7
## + B       1     312.33 12101 1618.1
## + ZN      1     150.36 12263 1624.8
## + CRIM    1     140.09 12273 1625.3
## + RAD     1      51.68 12362 1628.9
## <none>               12413 1629.0
## + INDUS   1      21.34 12392 1630.1
```

```
## + TAX    1     10.40 12403 1630.6
## + AGE    1      0.10 12413 1631.0
##
## Step:  AIC=1617.73
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS
##
##           Df Sum of Sq   RSS    AIC
## + B       1   273.644 11818 1608.2
## + ZN      1   162.963 11928 1612.9
## + CRIM    1   115.410 11976 1614.9
## + RAD     1    56.738 12035 1617.3
## <none>                 12091 1617.7
## + INDUS   1    31.075 12060 1618.4
## + TAX     1     4.175 12087 1619.5
## + AGE     1     1.808 12090 1619.7
##
## Step:  AIC=1608.17
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS + B
##
##           Df Sum of Sq   RSS    AIC
## + ZN      1   188.409 11629 1602.0
## + RAD     1   141.504 11676 1604.1
## + CRIM    1    54.898 11763 1607.8
## <none>                 11818 1608.2
## + INDUS   1    19.333 11798 1609.3
## + AGE     1     8.363 11809 1609.8
## + TAX     1     2.732 11815 1610.0
##
## Step:  AIC=1602.05
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS + B + ZN
##
##           Df Sum of Sq   RSS    AIC
## + CRIM    1    93.590 11536 1600.0
## + RAD     1    91.530 11538 1600.1
## <none>                 11629 1602.0
## + INDUS   1    19.788 11610 1603.2
## + TAX     1     3.859 11626 1603.9
## + AGE     1     1.098 11628 1604.0
##
## Step:  AIC=1599.97
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS + B + ZN + CRIM
##
##           Df Sum of Sq   RSS    AIC
## + RAD     1   224.324 11312 1592.0
## <none>                 11536 1600.0
## + INDUS   1    19.441 11516 1601.1
## + AGE     1     1.954 11534 1601.9
## + TAX     1     1.318 11534 1601.9
##
## Step:  AIC=1592.05
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS + B + ZN + CRIM +
##     RAD
##
##           Df Sum of Sq   RSS    AIC
```

```
## + TAX     1   267.996 11044 1581.9
## <none>               11312 1592.0
## + INDUS  1    38.798 11273 1592.3
## + AGE    1     0.024 11312 1594.0
##
## Step:  AIC=1581.94
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS + B + ZN + CRIM +
##     RAD + TAX
##
##          Df Sum of Sq   RSS    AIC
## <none>               11044 1581.9
## + INDUS  1   1.20733 11042 1583.9
## + AGE    1   0.15194 11043 1583.9
```

```r
AIC_backward = step(housing_boston_model, MEDV ~ 1, direction = "backward")
```

```
## Start:  AIC=1585.88
## MEDV ~ CRIM + ZN + INDUS + CHAS + NOX + RM + AGE + DIS + RAD +
##     TAX + PTRATIO + B + LSTAT
##
##           Df Sum of Sq   RSS    AIC
## - AGE      1      0.15 11042 1583.9
## - INDUS    1      1.20 11043 1583.9
## <none>                11042 1585.9
## - CHAS     1    216.39 11258 1593.7
## - TAX      1    230.55 11273 1594.3
## - CRIM     1    240.50 11283 1594.8
## - ZN       1    254.11 11296 1595.4
## - B        1    270.27 11312 1596.1
## - RAD      1    463.80 11506 1604.7
## - NOX      1    479.53 11522 1605.3
## - PTRATIO  1   1208.01 12250 1636.3
## - DIS      1   1238.20 12280 1637.5
## - RM       1   1854.49 12897 1662.3
## - LSTAT    1   2432.33 13474 1684.4
##
## Step:  AIC=1583.89
## MEDV ~ CRIM + ZN + INDUS + CHAS + NOX + RM + DIS + RAD + TAX +
##     PTRATIO + B + LSTAT
##
##           Df Sum of Sq   RSS    AIC
## - INDUS    1      1.21 11044 1581.9
## <none>                11042 1583.9
## - CHAS     1    217.54 11260 1591.7
## - TAX      1    230.41 11273 1592.3
## - CRIM     1    240.47 11283 1592.8
## - ZN       1    256.33 11299 1593.5
## - B        1    272.19 11314 1594.2
## - RAD      1    465.14 11507 1602.7
## - NOX      1    510.76 11553 1604.7
## - PTRATIO  1   1213.45 12256 1634.5
## - DIS      1   1361.30 12404 1640.6
## - RM       1   1944.98 12987 1663.8
## - LSTAT    1   2739.13 13781 1693.8
##
```

```
## Step:  AIC=1581.94
## MEDV ~ CRIM + ZN + CHAS + NOX + RM + DIS + RAD + TAX + PTRATIO +
##     B + LSTAT
##
##           Df Sum of Sq   RSS    AIC
## <none>                 11044 1581.9
## - CHAS     1    223.16 11267 1590.0
## - CRIM     1    242.00 11286 1590.9
## - ZN       1    255.54 11299 1591.5
## - TAX      1    268.00 11312 1592.0
## - B        1    271.28 11315 1592.2
## - RAD      1    491.00 11534 1601.9
## - NOX      1    538.08 11582 1604.0
## - PTRATIO  1   1223.51 12267 1633.0
## - DIS      1   1450.85 12494 1642.3
## - RM       1   1953.52 12997 1662.2
## - LSTAT    1   2746.90 13790 1692.1
```

```r
AIC_forward_score = AIC(AIC_forward)
AIC_backward_score = AIC(AIC_backward)

AIC_forward_score
```

```
## [1] 3017.072
```

```r
AIC_backward_score
```

```
## [1] 3017.072
```

The forward model and backward model are the same.

(2) Do forward and Backward model selection using BIC. Are the model obtained from forward and backward selection are same?

```r
# Find n
n = dim(housing_boston)[2]-1
n
```

```
## [1] 13
```

```r
BIC_forward = step(lm(MEDV ~ 1,data = housing_boston),lm_object,
                   direction = "forward",k = log(n))
```

```
## Start:  AIC=2243.62
## MEDV ~ 1
##
##           Df Sum of Sq   RSS    AIC
## + LSTAT    1   23275.8 19438 1848.6
## + RM       1   20653.6 22060 1912.5
## + PTRATIO  1   11040.8 31673 2095.2
## + INDUS    1   10011.2 32703 2111.3
## + TAX      1    9377.1 33337 2121.0
## + NOX      1    7798.8 34915 2144.4
## + CRIM     1    6438.6 36276 2163.7
## + RAD      1    6222.9 36491 2166.7
## + AGE      1    6068.7 36645 2168.8
## + ZN       1    5547.9 37166 2175.9
```

```
## + B         1    4747.7 37966 2186.7
## + DIS       1    2667.4 40047 2213.6
## + CHAS      1    1313.6 41401 2230.4
## <none>                  42714 2243.6
##
## Step:  AIC=1848.6
## MEDV ~ LSTAT
##
##           Df Sum of Sq   RSS    AIC
## + RM       1    4023.5 15415 1734.0
## + PTRATIO  1    2707.8 16730 1775.4
## + CHAS     1     781.7 18657 1830.4
## + DIS      1     779.3 18659 1830.5
## + AGE      1     310.5 19128 1843.0
## + TAX      1     275.2 19163 1844.0
## + B        1     198.7 19240 1846.0
## + ZN       1     159.2 19279 1847.0
## + CRIM     1     146.4 19292 1847.3
## + INDUS    1     103.1 19335 1848.5
## <none>                 19438 1848.6
## + RAD      1      26.5 19412 1850.5
## + NOX      1       5.5 19433 1851.0
##
## Step:  AIC=1734.05
## MEDV ~ LSTAT + RM
##
##           Df Sum of Sq   RSS    AIC
## + PTRATIO  1   1738.57 13676 1676.2
## + CHAS     1    545.44 14869 1718.4
## + B        1    512.40 14902 1719.5
## + TAX      1    425.87 14989 1722.5
## + DIS      1    355.52 15059 1724.8
## + CRIM     1    310.60 15104 1726.3
## + RAD      1    183.44 15231 1730.6
## <none>                 15415 1734.0
## + INDUS    1     64.00 15351 1734.5
## + ZN       1     56.06 15359 1734.8
## + AGE      1     21.70 15393 1735.9
## + NOX      1     13.84 15401 1736.2
##
## Step:  AIC=1676.18
## MEDV ~ LSTAT + RM + PTRATIO
##
##           Df Sum of Sq   RSS    AIC
## + DIS      1    508.25 13168 1659.6
## + B        1    388.72 13288 1664.2
## + CHAS     1    372.75 13304 1664.8
## + CRIM     1    120.37 13556 1674.3
## + AGE      1     70.86 13605 1676.1
## <none>                 13676 1676.2
## + TAX      1     42.82 13633 1677.2
## + NOX      1     22.92 13653 1677.9
## + ZN       1     16.23 13660 1678.2
## + RAD      1      6.03 13670 1678.5
```

```
## + INDUS  1       0.53 13676 1678.7
##
## Step:  AIC=1659.62
## MEDV ~ LSTAT + RM + PTRATIO + DIS
##
##           Df Sum of Sq   RSS    AIC
## + NOX    1     754.60 12413 1632.4
## + B      1     502.53 12665 1642.5
## + CHAS   1     261.62 12906 1652.0
## + INDUS  1     256.19 12912 1652.3
## + TAX    1     238.61 12929 1653.0
## + CRIM   1     231.41 12937 1653.2
## + ZN     1     143.45 13024 1656.7
## <none>              13168 1659.6
## + AGE    1      58.18 13110 1660.0
## + RAD    1      23.17 13145 1661.3
##
## Step:  AIC=1632.39
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX
##
##           Df Sum of Sq   RSS    AIC
## + CHAS   1     321.93 12091 1621.7
## + B      1     312.33 12101 1622.1
## + ZN     1     150.36 12263 1628.8
## + CRIM   1     140.09 12273 1629.2
## <none>              12413 1632.4
## + RAD    1      51.68 12362 1632.8
## + INDUS  1      21.34 12392 1634.1
## + TAX    1      10.40 12403 1634.5
## + AGE    1       0.10 12413 1635.0
##
## Step:  AIC=1621.68
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS
##
##           Df Sum of Sq   RSS    AIC
## + B      1    273.644 11818 1612.7
## + ZN     1    162.963 11928 1617.4
## + CRIM   1    115.410 11976 1619.4
## <none>              12091 1621.7
## + RAD    1     56.738 12035 1621.9
## + INDUS  1     31.075 12060 1623.0
## + TAX    1      4.175 12087 1624.1
## + AGE    1      1.808 12090 1624.2
##
## Step:  AIC=1612.69
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS + B
##
##           Df Sum of Sq   RSS    AIC
## + ZN     1    188.409 11629 1607.1
## + RAD    1    141.504 11676 1609.2
## <none>              11818 1612.7
## + CRIM   1     54.898 11763 1612.9
## + INDUS  1     19.333 11798 1614.4
## + AGE    1      8.363 11809 1614.9
```

```
## + TAX    1      2.732 11815 1615.1
##
## Step:  AIC=1607.13
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS + B + ZN
##
##           Df Sum of Sq   RSS    AIC
## + CRIM   1     93.590 11536 1605.6
## + RAD    1     91.530 11538 1605.7
## <none>               11629 1607.1
## + INDUS  1     19.788 11610 1608.8
## + TAX    1      3.859 11626 1609.5
## + AGE    1      1.098 11628 1609.7
##
## Step:  AIC=1605.62
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS + B + ZN + CRIM
##
##           Df Sum of Sq   RSS    AIC
## + RAD    1    224.324 11312 1598.3
## <none>               11536 1605.6
## + INDUS  1     19.441 11516 1607.3
## + AGE    1      1.954 11534 1608.1
## + TAX    1      1.318 11534 1608.1
##
## Step:  AIC=1598.27
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS + B + ZN + CRIM +
##      RAD
##
##           Df Sum of Sq   RSS    AIC
## + TAX    1    267.996 11044 1588.7
## <none>               11312 1598.3
## + INDUS  1     38.798 11273 1599.1
## + AGE    1      0.024 11312 1600.8
##
## Step:  AIC=1588.72
## MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS + B + ZN + CRIM +
##      RAD + TAX
##
##           Df Sum of Sq   RSS    AIC
## <none>               11044 1588.7
## + INDUS  1    1.20733 11042 1591.2
## + AGE    1    0.15194 11043 1591.3
```

```r
BIC_backward = step(housing_boston_model, MEDV ~ 1 , direction="backward", k=log(n))
```

```
## Start:  AIC=1593.79
## MEDV ~ CRIM + ZN + INDUS + CHAS + NOX + RM + AGE + DIS + RAD +
##      TAX + PTRATIO + B + LSTAT
##
##            Df Sum of Sq   RSS    AIC
## - AGE      1      0.15 11042 1591.2
## - INDUS    1      1.20 11043 1591.3
## <none>               11042 1593.8
## - CHAS     1    216.39 11258 1601.0
## - TAX      1    230.55 11273 1601.7
## - CRIM     1    240.50 11283 1602.1
```

```
## - ZN         1    254.11 11296 1602.7
## - B          1    270.27 11312 1603.4
## - RAD        1    463.80 11506 1612.0
## - NOX        1    479.53 11522 1612.7
## - PTRATIO    1   1208.01 12250 1643.7
## - DIS        1   1238.20 12280 1644.9
## - RM         1   1854.49 12897 1669.6
## - LSTAT      1   2432.33 13474 1691.8
##
## Step:  AIC=1591.23
## MEDV ~ CRIM + ZN + INDUS + CHAS + NOX + RM + DIS + RAD + TAX +
##     PTRATIO + B + LSTAT
##
##            Df Sum of Sq   RSS    AIC
## - INDUS     1      1.21 11044 1588.7
## <none>                  11042 1591.2
## - CHAS      1    217.54 11260 1598.5
## - TAX       1    230.41 11273 1599.1
## - CRIM      1    240.47 11283 1599.5
## - ZN        1    256.33 11299 1600.3
## - B         1    272.19 11314 1601.0
## - RAD       1    465.14 11507 1609.5
## - NOX       1    510.76 11553 1611.5
## - PTRATIO   1   1213.45 12256 1641.3
## - DIS       1   1361.30 12404 1647.4
## - RM        1   1944.98 12987 1670.6
## - LSTAT     1   2739.13 13781 1700.6
##
## Step:  AIC=1588.72
## MEDV ~ CRIM + ZN + CHAS + NOX + RM + DIS + RAD + TAX + PTRATIO +
##     B + LSTAT
##
##            Df Sum of Sq   RSS    AIC
## <none>                  11044 1588.7
## - CHAS      1    223.16 11267 1596.3
## - CRIM      1    242.00 11286 1597.1
## - ZN        1    255.54 11299 1597.7
## - TAX       1    268.00 11312 1598.3
## - B         1    271.28 11315 1598.4
## - RAD       1    491.00 11534 1608.1
## - NOX       1    538.08 11582 1610.2
## - PTRATIO   1   1223.51 12267 1639.2
## - DIS       1   1450.85 12494 1648.5
## - RM        1   1953.52 12997 1668.4
## - LSTAT     1   2746.90 13790 1698.3
```

```r
BIC_forward_score = BIC(BIC_forward)
BIC_backward_score = BIC(BIC_backward)
BIC_forward
```

```
##
## Call:
## lm(formula = MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS +
##     B + ZN + CRIM + RAD + TAX, data = housing_boston)
##
```

```
## Coefficients:
## (Intercept)         LSTAT           RM       PTRATIO           DIS
##   36.535739     -0.525325     3.792330     -0.954188     -1.493706
##         NOX          CHAS            B            ZN          CRIM
##  -17.315910      2.694981     0.009299      0.045644     -0.107681
##         RAD           TAX
##    0.296801     -0.011661
```

BIC_backward

```
##
## Call:
## lm(formula = MEDV ~ CRIM + ZN + CHAS + NOX + RM + DIS + RAD +
##     TAX + PTRATIO + B + LSTAT, data = housing_boston)
##
## Coefficients:
## (Intercept)          CRIM            ZN          CHAS           NOX
##   36.535739     -0.107681     0.045644      2.694981     -17.315910
##          RM           DIS           RAD           TAX       PTRATIO
##    3.792330     -1.493706     0.296801     -0.011661     -0.954188
##           B         LSTAT
##    0.009299     -0.525325
```

The forward model and backward model are the same.

(3) Compare the forward model from Q1 and Q2.

AIC_forward

```
##
## Call:
## lm(formula = MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS +
##     B + ZN + CRIM + RAD + TAX, data = housing_boston)
##
## Coefficients:
## (Intercept)         LSTAT           RM       PTRATIO           DIS
##   36.535739     -0.525325     3.792330     -0.954188     -1.493706
##         NOX          CHAS            B            ZN          CRIM
##  -17.315910      2.694981     0.009299      0.045644     -0.107681
##         RAD           TAX
##    0.296801     -0.011661
```

BIC_forward

```
##
## Call:
## lm(formula = MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS +
##     B + ZN + CRIM + RAD + TAX, data = housing_boston)
##
## Coefficients:
## (Intercept)         LSTAT           RM       PTRATIO           DIS
##   36.535739     -0.525325     3.792330     -0.954188     -1.493706
##         NOX          CHAS            B            ZN          CRIM
##  -17.315910      2.694981     0.009299      0.045644     -0.107681
##         RAD           TAX
##    0.296801     -0.011661
```

Therefore, we can see that the forward model from Q1 and Q2 are the same.

(4) Compare the backward models from Q1 and Q2.

`AIC_backward`

```
##
## Call:
## lm(formula = MEDV ~ CRIM + ZN + CHAS + NOX + RM + DIS + RAD +
##     TAX + PTRATIO + B + LSTAT, data = housing_boston)
##
## Coefficients:
## (Intercept)          CRIM            ZN          CHAS           NOX
##   36.535739     -0.107681      0.045644      2.694981    -17.315910
##          RM           DIS           RAD           TAX       PTRATIO
##    3.792330     -1.493706      0.296801     -0.011661     -0.954188
##           B         LSTAT
##    0.009299     -0.525325
```

`BIC_backward`

```
##
## Call:
## lm(formula = MEDV ~ CRIM + ZN + CHAS + NOX + RM + DIS + RAD +
##     TAX + PTRATIO + B + LSTAT, data = housing_boston)
##
## Coefficients:
## (Intercept)          CRIM            ZN          CHAS           NOX
##   36.535739     -0.107681      0.045644      2.694981    -17.315910
##          RM           DIS           RAD           TAX       PTRATIO
##    3.792330     -1.493706      0.296801     -0.011661     -0.954188
##           B         LSTAT
##    0.009299     -0.525325
```

Therefore, we can see that the backward model from Q1 and Q2 are the same.

(5) See if all variables obtained in the forward AIC model is significant or not ?

`AIC_forward`

```
##
## Call:
## lm(formula = MEDV ~ LSTAT + RM + PTRATIO + DIS + NOX + CHAS +
##     B + ZN + CRIM + RAD + TAX, data = housing_boston)
##
## Coefficients:
## (Intercept)         LSTAT            RM       PTRATIO           DIS
##   36.535739     -0.525325      3.792330     -0.954188     -1.493706
##         NOX          CHAS             B            ZN          CRIM
##  -17.315910      2.694981      0.009299      0.045644     -0.107681
##         RAD           TAX
##    0.296801     -0.011661
```

`summary(housing_boston_model)`

```
##
```

```
## Call:
## lm(formula = lm_object, data = housing_boston)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -15.5642  -2.7248  -0.5312   1.7687  26.1511
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  36.634820   5.102042   7.180 2.59e-12 ***
## CRIM         -0.107417   0.032847  -3.270 0.001150 **
## ZN            0.046121   0.013721   3.361 0.000836 ***
## INDUS         0.014270   0.061653   0.231 0.817053
## CHAS          2.671108   0.861116   3.102 0.002033 **
## NOX         -17.633543   3.818718  -4.618 4.96e-06 ***
## RM            3.794304   0.417836   9.081  < 2e-16 ***
## AGE           0.001076   0.013205   0.082 0.935062
## DIS          -1.479170   0.199347  -7.420 5.19e-13 ***
## RAD           0.301535   0.066398   4.541 7.04e-06 ***
## TAX          -0.012054   0.003765  -3.202 0.001454 **
## PTRATIO      -0.958871   0.130831  -7.329 9.60e-13 ***
## B             0.009305   0.002684   3.467 0.000573 ***
## LSTAT        -0.527600   0.050732 -10.400  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.742 on 491 degrees of freedom
## Multiple R-squared:  0.7415, Adjusted R-squared:  0.7346
## F-statistic: 108.3 on 13 and 491 DF,  p-value: < 2.2e-16
```

Yes, all variables obtained in the forward AIC model are significant.

6.Get a 95% confidence interval for the variables for forward AIC model.

```
housing_boston_model = lm(formula = MEDV~LSTAT+RM+PTRATIO+DIS+NOX+CHAS+B+ZN+
                          CRIM+RAD+TAX, data = housing_boston)
confint(housing_boston_model, c("LSTAT", "RM", "PTRATIO", "DIS", "NOX","CHAS","B","ZN","CRIM","RAD","TA
```

```
##                  2.5 %        97.5 %
## LSTAT     -0.618532581  -0.432117084
## RM         2.994441207   4.590219465
## PTRATIO   -1.207861600  -0.700513725
## DIS       -1.858375283  -1.129037143
## NOX      -24.257638106 -10.374180978
## CHAS       1.017370620   4.372591600
## B          0.004048764   0.014548759
## ZN         0.019091743   0.072196537
## CRIM      -0.172050348  -0.043312125
## RAD        0.172243649   0.421357523
## TAX       -0.018284192  -0.005036851
```

Therefore, the 95% confidence interval for the variables for forward AIC model is shown above.