multi\_reg\_1.R

SANGHOOJEFFREY

Tue Jun 26 19:45:04 2018

if(!require(car)) install.packages("car"); library(car)

## Loading required package: car

## Warning: package 'car' was built under R version 3.4.4

## Loading required package: carData

## Warning: package 'carData' was built under R version 3.4.4

# 다중회귀모형  
# 단순회귀모형은 반응변수와 설명변수가 각각 1개인 경우  
# 설명변수가 여러개와 반응변수 간 선형관계식을 세우는 분석이 다중회귀분석  
par(mfrow=c(1,1))  
  
data(iris)  
head(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa

# Sepal.Length = Sepal.width + Petal.length+Petal.width + species 로 구성된 회귀식을 세워보자.   
# 회귀분석 전 설명변수와 반응변수의 관계를 시각적으로 확인하기 위한 산점도를 그려보자.  
# 여러 변수 간 상관성을 살펴보기 위해 pairs.panels() 함수 이용  
  
if (!require(psych)) install.packages("psych"); library(psych)

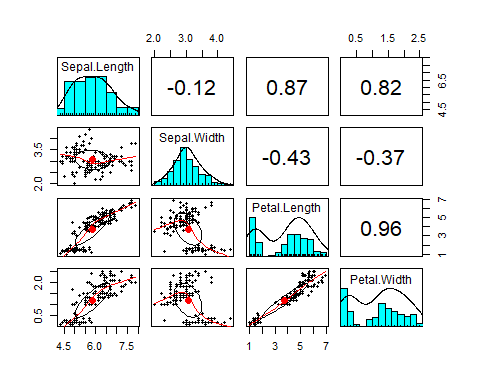
## Loading required package: psych

## Warning: package 'psych' was built under R version 3.4.4

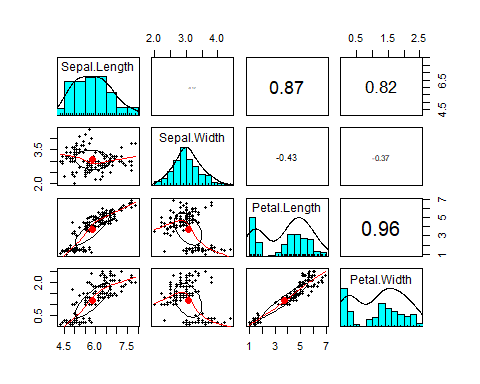
##   
## Attaching package: 'psych'

## The following object is masked from 'package:car':  
##   
## logit

pairs.panels(iris[,1:4], scale=FALSE)



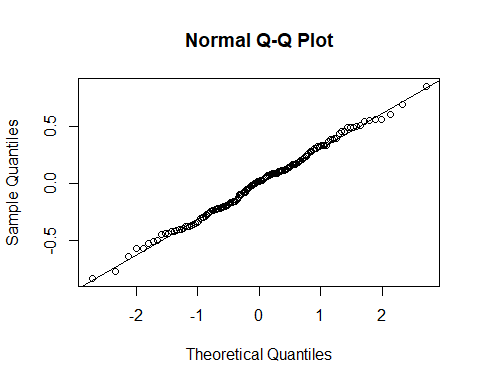
pairs.panels(iris[,1:4], scale=TRUE)



out.i <- lm(Sepal.Length ~ Sepal.Width + Petal.Length+ Petal.Width, data=iris)  
summary(out.i)

##   
## Call:  
## lm(formula = Sepal.Length ~ Sepal.Width + Petal.Length + Petal.Width,   
## data = iris)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.82816 -0.21989 0.01875 0.19709 0.84570   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.85600 0.25078 7.401 9.85e-12 \*\*\*  
## Sepal.Width 0.65084 0.06665 9.765 < 2e-16 \*\*\*  
## Petal.Length 0.70913 0.05672 12.502 < 2e-16 \*\*\*  
## Petal.Width -0.55648 0.12755 -4.363 2.41e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3145 on 146 degrees of freedom  
## Multiple R-squared: 0.8586, Adjusted R-squared: 0.8557   
## F-statistic: 295.5 on 3 and 146 DF, p-value: < 2.2e-16

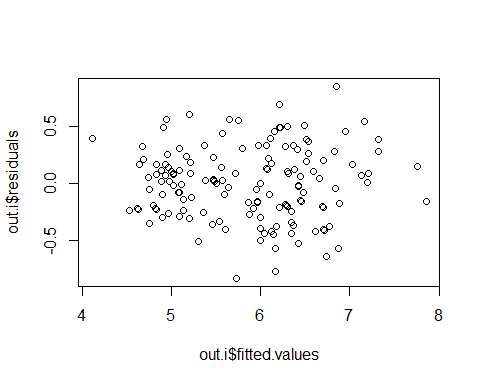
# 다중회귀모형의 경우 설명변수가 증가할 수록 R^2가 증가한다.  
# 불필요한 독립변수가 모형에 반영될 필요가 없으므로 독립변수의 수에 벌점을 준 수정된 R^2로 최적모형 결정한다.  
# adj. R^2 = 1 - (SSE/(n-k-1)) / (SST/(n-1))  
# 여기서 n은 데이터의 수, k는 독립 변수의 수이다.   
  
# 정규성 검정  
qqnorm(out.i$residuals)  
qqline(out.i$residuals)



shapiro.test(out.i$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: out.i$residuals  
## W = 0.99559, p-value = 0.9349

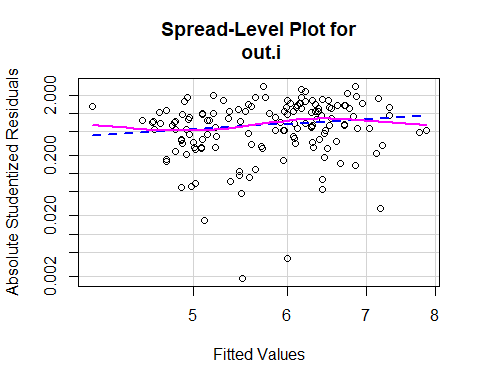
# 독립성 검정  
plot(out.i$fitted.values, out.i$residuals)



durbinWatsonTest(out.i)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.03992126 2.060382 0.784  
## Alternative hypothesis: rho != 0

# 등분산성 검정  
spreadLevelPlot(out.i)



##   
## Suggested power transformation: -0.1805994

ncvTest(out.i)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 4.448612 Df = 1 p = 0.03492962

# 다중공선성 문제, 독립변수들 간 높은 상관관계가 있는 경우 다중회귀모형 추정이 어려움  
# 확인하기 위해 분산팽창지수 (vif)를 확인해야함.  
# 만약 vif>5이상이면 변수선택, 능형회귀모형, 주성분회귀모형 등을 통해 다중공선성 문제 해결이 필요  
  
vif(out.i) # Petal.Length와 Petal.Width의 vif>5이상임

## Sepal.Width Petal.Length Petal.Width   
## 1.270815 15.097572 14.234335

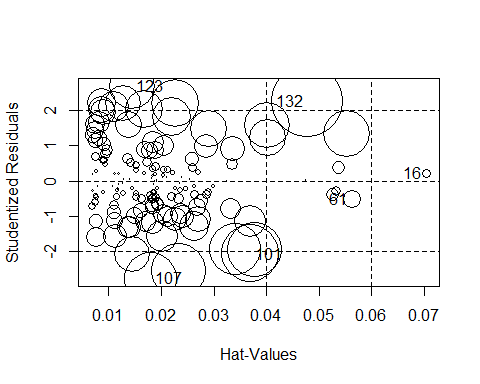
out.i2 <- lm(Sepal.Length ~ Sepal.Width + Petal.Width, data=iris) # 높은 vif 순으로 제거  
vif(out.i2) # 다중공선성 문제 해결

## Sepal.Width Petal.Width   
## 1.154799 1.154799

summary(out.i2)

##   
## Call:  
## lm(formula = Sepal.Length ~ Sepal.Width + Petal.Width, data = iris)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2076 -0.2288 -0.0450 0.2266 1.1810   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.45733 0.30919 11.18 < 2e-16 \*\*\*  
## Sepal.Width 0.39907 0.09111 4.38 2.24e-05 \*\*\*  
## Petal.Width 0.97213 0.05210 18.66 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4511 on 147 degrees of freedom  
## Multiple R-squared: 0.7072, Adjusted R-squared: 0.7033   
## F-statistic: 177.6 on 2 and 147 DF, p-value: < 2.2e-16

# 회귀모형의 진단 및 보정  
# 지렛대점, 영향점, 이상점을 알아보자.  
influencePlot(out.i2)



## StudRes Hat CookD  
## 16 0.2243801 0.07040119 0.001279224  
## 61 -0.5181403 0.05640452 0.005376110  
## 101 -2.0661597 0.03702741 0.053525880  
## 107 -2.7617060 0.01791588 0.044378518  
## 123 2.6926993 0.01440015 0.033871613  
## 132 2.2619187 0.04787655 0.083419653

# StudRes : 이상점 여부 확인  
# Hat : 모자행렬을 이용한 지렛대 점 확인  
# CookD : 영향점 확인  
  
# 이상점은 outlierTest()함수를 통해 확인  
# 이상점이 많을 수록 모형의 결정계수인 R^2 가 감소됨  
outlierTest(out.i2)

## No Studentized residuals with Bonferonni p < 0.05  
## Largest |rstudent|:  
## rstudent unadjusted p-value Bonferonni p  
## 107 -2.761706 0.0064892 0.97338

# Hat이 2(k+1)/n 이면 지렛대 점으로 주의깊게 볼 필요가 있음  
2\*(2+1)/nrow(iris)

## [1] 0.04

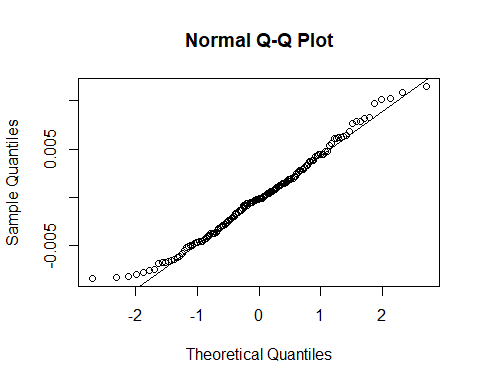
# 쿡의 거리는 4/(n-k-1)보다 크면 영향점으로 판단   
4/(nrow(iris)-2-1)

## [1] 0.02721088

# 영향점 : 101, 107, 123, 132  
# 지렛대점 : 16, 61, 132  
# 이상점 : 107   
  
# 영향점과 이상점은 회귀모형에 안 좋은 영향을 미치는 자료로 제거시 더 좋은 회귀모형식을 얻을 수 있다.  
# 단, 두 개 이상의 이상점이 이웃하면 서로의 이상점 효과가 상쇄될 수 있다.   
# 모형이 변하면 이상점을 다시 조사해야 한다.  
# 데이터셋이 클 경우 한 두개의 이상점은 문제되지 않지만,  
# 이상점이 그룹을 형성하는 경우에는 분석에 주의해야 한다.  
  
# 이 자료의 경우 101, 107, 123, 132가 회귀모형에 악영향을 미친다고 판단되므로 제거하여 회귀모형을 다시 세우면  
  
iris2 <- iris[c(-101,-107, -123, -132),]  
  
out.i3 <- lm(Sepal.Length^-2 ~ Sepal.Width + Petal.Width, data=iris2)   
summary(out.i3)

##   
## Call:  
## lm(formula = Sepal.Length^-2 ~ Sepal.Width + Petal.Width, data = iris2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0083568 -0.0033739 -0.0001687 0.0028311 0.0114947   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0543934 0.0031708 17.154 < 2e-16 \*\*\*  
## Sepal.Width -0.0035895 0.0009328 -3.848 0.000179 \*\*\*  
## Petal.Width -0.0103619 0.0005343 -19.393 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.004492 on 143 degrees of freedom  
## Multiple R-squared: 0.7328, Adjusted R-squared: 0.7291   
## F-statistic: 196.1 on 2 and 143 DF, p-value: < 2.2e-16

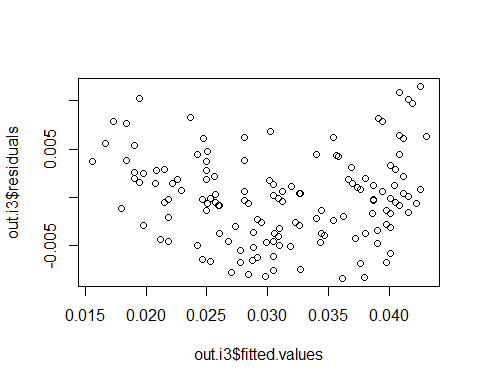
# 정규성 검정  
qqnorm(out.i3$residuals)  
qqline(out.i3$residuals)



shapiro.test(out.i3$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: out.i3$residuals  
## W = 0.98497, p-value = 0.1126

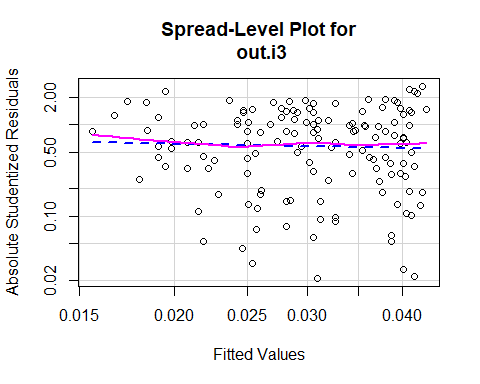
# 독립성 검정  
plot(out.i3$fitted.values, out.i3$residuals)



durbinWatsonTest(out.i3)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.1525169 1.689486 0.058  
## Alternative hypothesis: rho != 0

# 등분산성 검정  
spreadLevelPlot(out.i3)

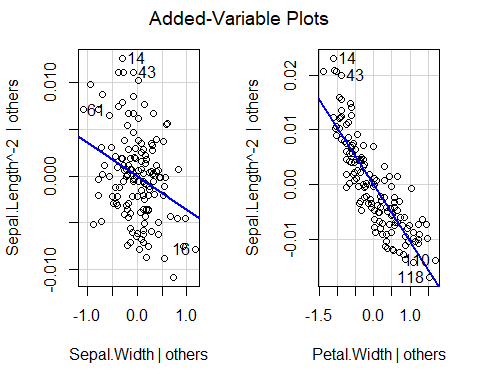


##   
## Suggested power transformation: 1.130383

ncvTest(out.i3)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 0.8141326 Df = 1 p = 0.3669013

# 추가그림은 반응변수와 설명변수의 관계를 2차식으로 표현한 그림이다.  
avPlots(out.i3)



# 다중공선선, 영향점, 이상점, 지렛대점을 확인  
  
  
#########################################################################  
summary(iris) # Species 는 범주형 변수임

## Sepal.Length Sepal.Width Petal.Length Petal.Width   
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100   
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300   
## Median :5.800 Median :3.000 Median :4.350 Median :1.300   
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199   
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800   
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500   
## Species   
## setosa :50   
## versicolor:50   
## virginica :50   
##   
##   
##

# 범주형 자료는 factor로 변환하여 분석   
# factor로 변환시 가변수로 처리하여 회귀모형을 세움  
  
out.i2 <- lm(Sepal.Length ~ Sepal.Width + Petal.Length+ Petal.Width+as.factor(Species), data=iris)  
summary(out.i2)

##   
## Call:  
## lm(formula = Sepal.Length ~ Sepal.Width + Petal.Length + Petal.Width +   
## as.factor(Species), data = iris)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.79424 -0.21874 0.00899 0.20255 0.73103   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.17127 0.27979 7.760 1.43e-12 \*\*\*  
## Sepal.Width 0.49589 0.08607 5.761 4.87e-08 \*\*\*  
## Petal.Length 0.82924 0.06853 12.101 < 2e-16 \*\*\*  
## Petal.Width -0.31516 0.15120 -2.084 0.03889 \*   
## as.factor(Species)versicolor -0.72356 0.24017 -3.013 0.00306 \*\*   
## as.factor(Species)virginica -1.02350 0.33373 -3.067 0.00258 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3068 on 144 degrees of freedom  
## Multiple R-squared: 0.8673, Adjusted R-squared: 0.8627   
## F-statistic: 188.3 on 5 and 144 DF, p-value: < 2.2e-16

# 범주형 결과에 대한 해석  
# default : setosa  
# versicolor : setosa종에 비해 versicolor종의 Sepal.Length가 -0.72356 작다  
# virginica : setosa종에 비해 virginica종의 Sepal.Length가 -1.02350 작다  
  
# 아이리스의 종별 Sepal.Length의 식을 다시 정리하면 아래와 같다.  
# setosa : 2.17+Sepal.width\*0.49 + petal.length\*0.82+Petal.width\*-0.31  
# versicolor : 2.17 -0.72356 +Sepal.width\*0.49 + petal.length\*0.82+Petal.width\*-0.31  
# virginica : 2.17 - 1.02350 +Sepal.width\*0.49 + petal.length\*0.82+Petal.width\*-0.31  
  
##########################################################################  
# 회귀모형의 설명변수가 많은 경우 중요한 설명변수를 선택해야한다.  
# 1. 전진선택법  
# 2. 후진제거법  
# 3. 단계적 방법  
# 위의 세가지 방법은 step()함수를 통해 진행할 수 있다.  
  
if(!require(mlbench)) install.packages("mlbench"); library(mlbench)

## Loading required package: mlbench

data("BostonHousing")  
  
m <- lm (medv ~ ., data=BostonHousing) # 여기서 .은 medv를 제외한 모든 변수를 설명변수로 간  
m2 <- step(m, ddirection="forward") # 전진선택법

## Start: AIC=1589.64  
## medv ~ crim + zn + indus + chas + nox + rm + age + dis + rad +   
## tax + ptratio + b + 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  
## - b 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 + b + 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  
## - b 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 +   
## b + 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  
## - b 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

m3 <- step(m, ddirection="bacward") # 후진제거법

## Start: AIC=1589.64  
## medv ~ crim + zn + indus + chas + nox + rm + age + dis + rad +   
## tax + ptratio + b + 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  
## - b 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 + b + 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  
## - b 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 +   
## b + 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  
## - b 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

m4 <- step(m, ddirection="both") # 단계적방법

## Start: AIC=1589.64  
## medv ~ crim + zn + indus + chas + nox + rm + age + dis + rad +   
## tax + ptratio + b + 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  
## - b 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 + b + 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  
## - b 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 +   
## b + 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  
## - b 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

formula(m2)

## medv ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +   
## b + lstat

formula(m3)

## medv ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +   
## b + lstat

formula(m4)

## medv ~ crim + zn + chas + nox + rm + dis + rad + tax + ptratio +   
## b + lstat

# 모든 가능한 경우를 고려한 최적모형 찾기  
# leaps::regsubsets()함수는 2N개의 회귀 모형을 만들어 비교를 수행하는 함수  
  
if(!require(leaps)) install.packages("leaps"); library(leaps)

## Loading required package: leaps

m5 <- regsubsets(medv ~., data=BostonHousing)  
summary(m5)

## Subset selection object  
## Call: regsubsets.formula(medv ~ ., data = BostonHousing)  
## 13 Variables (and intercept)  
## Forced in Forced out  
## crim FALSE FALSE  
## zn FALSE FALSE  
## indus FALSE FALSE  
## chas1 FALSE FALSE  
## nox FALSE FALSE  
## rm FALSE FALSE  
## age FALSE FALSE  
## dis FALSE FALSE  
## rad FALSE FALSE  
## tax FALSE FALSE  
## ptratio FALSE FALSE  
## b FALSE FALSE  
## lstat FALSE FALSE  
## 1 subsets of each size up to 8  
## Selection Algorithm: exhaustive  
## crim zn indus chas1 nox rm age dis rad tax ptratio b lstat  
## 1 ( 1 ) " " " " " " " " " " " " " " " " " " " " " " " " "\*"   
## 2 ( 1 ) " " " " " " " " " " "\*" " " " " " " " " " " " " "\*"   
## 3 ( 1 ) " " " " " " " " " " "\*" " " " " " " " " "\*" " " "\*"   
## 4 ( 1 ) " " " " " " " " " " "\*" " " "\*" " " " " "\*" " " "\*"   
## 5 ( 1 ) " " " " " " " " "\*" "\*" " " "\*" " " " " "\*" " " "\*"   
## 6 ( 1 ) " " " " " " "\*" "\*" "\*" " " "\*" " " " " "\*" " " "\*"   
## 7 ( 1 ) " " " " " " "\*" "\*" "\*" " " "\*" " " " " "\*" "\*" "\*"   
## 8 ( 1 ) " " "\*" " " "\*" "\*" "\*" " " "\*" " " " " "\*" "\*" "\*"

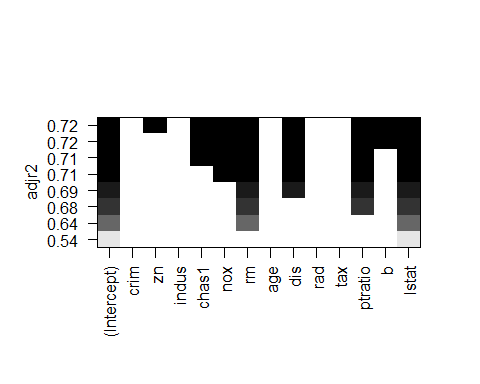
summary(m5)$bic

## [1] -385.0521 -496.2582 -549.4767 -561.9884 -585.6823 -592.9553 -598.2295  
## [8] -600.1663

summary(m5)$adjr2

## [1] 0.5432418 0.6371245 0.6767036 0.6878351 0.7051702 0.7123567 0.7182560  
## [8] 0.7222072

plot(m5, scale="adjr2")



plot(m5, scale="bic")

