**Homework 8 by Haritha Pulletikurti**

**Question 11.1**

Using the crime data set uscrime.txt from Questions 8.2, 9.1, and 10.1, build a regression model using:

# Stepwise regression

1. Lasso
2. Elastic net

For Parts 2 and 3, remember to scale the data first – otherwise, the regression coefficients will be on different scales and the constraint won’t have the desired effect.

For Parts 2 and 3, use the glmnet function in R.

Notes on R:

* For the elastic net model, what we called λ in the videos, glmnet calls “alpha”; you can get a range of results by varying alpha from 1 (lasso) to 0 (ridge regression) [and, of course, other values of alpha in between].
* In a function call like glmnet(x,y,family=”mgaussian”,alpha=1) the predictors x need to be in R’s matrix format, rather than data frame format. You can convert a data frame to a matrix using as.matrix – for example, x <- as.matrix(data[,1:n-1])
* Rather than specifying a value of T, glmnet returns models for a variety of values of T.

Answer:

**Stepwise Regression Techniques:**  As we all know,

# Bias: is the difference between average prediction of our model and the correct value we are trying to predict.

# Variance: is the variability of the model prediction based on different data sets i.e. Training / validation/Test data sets.

**Best Model:** We all know the best model is considered to have low bias and low variance.

If n = Number of data points and p = number of predictors

Case 1: if n is very large than p i.e. n >> p, then the least squared regression model tends to have low bias and low variance.

# Case 2: if n is not much larger than p i.e. n >p, then there can be lot of variability in least squares fit resulting in overfitting.

Case 3: If n is smaller than p, i.e. n < p, then the variance is infinite as there is no longer a unique least squared estimate.

There are many approaches for variable selection i.e. excluding the irrelevant variables from a multiple regression model like the ones mentioned below

**Forward Step wise Regression:**

Step 1: Let Mo denote the null model which contains no predictors.

# Step 2: For k = 0, …, p-1

1. Consider all p-k models that augment the predictors in the set Mk with one additional predictor.
2. Choose best among the p-k models, call it Mk+1. Here best is defined as having the smallest RSS or highest R^2.

Step3: Select the single best model from among the Mo, …, Mp using cross validated predictor error AIC, BIC or adjusted R^2.

**Backward Stepwise Elimination:**

Step 1: Let Mo denote the full model which contains all the predictors.

# Step 2: For k = p,p-1, …,1

1. Consider all k models that contain all but one of the predictors in the set Mk, for a total of k-1 predictors.
2. Choose best among these k models, call it Mk-1. Here best is defined as having the smallest RSS or highest R^2.

Step3: Select the single best model from among the Mo, …, Mp using cross validated predictor error AIC, BIC or adjusted R^2.

Backward selection requires that the number of samples n is larger than number of variables in p so that full model can be fit. In contrast Forward selection model can be used even when n < p and so is the only viable subset method when p is large.

# Stepwise Regression: is a combination of both forward selection and backward elimination.

# 1.Start with no predictors.

2.1 Find the best new predictor if it is good enough

a) add that factor, fit model with current set of factors.

b) Remove factors with high p value.

# c) Check if we have enough factors. If not repeat from steps 2.1.

# 2.2 If the chosen new predictor is not good enough

a) remove factors with high p-value

b) fit the model with final set of factors.

**Optimal Model:** The model that contains all the predictors will always have the smallest RSS and the largest R^2, since these quantities are related to training error. We should choose the model that has the low-test error. Since these measurements are not suitable to choose between models with different number of predictors, the best way to choose the model is by checking AIC, BIC and Adjusted R^2.

[References: An Introduction to Statistical Learning Text book].

**Forward Regression**  
  
rm(list = ls())  
set.seed(82)  
uscrime<- read.table("uscrime.txt", stringsAsFactors = FALSE, header = TRUE)  
  
uscrime[1:3,]

## M So Ed Po1 Po2 LF M.F Pop NW U1 U2 Wealth Ineq Prob  
## 1 15.1 1 9.1 5.8 5.6 0.510 95.0 33 30.1 0.108 4.1 3940 26.1 0.084602  
## 2 14.3 0 11.3 10.3 9.5 0.583 101.2 13 10.2 0.096 3.6 5570 19.4 0.029599  
## 3 14.2 1 8.9 4.5 4.4 0.533 96.9 18 21.9 0.094 3.3 3180 25.0 0.083401  
## Time Crime  
## 1 26.2011 791  
## 2 25.2999 1635  
## 3 24.3006 578

# Scale the data  
Scaleduscrime <- as.data.frame(scale(uscrime[,c(1,3:15)]))  
Scaleduscrime <- cbind(uscrime[,2],Scaleduscrime,uscrime[,16])  
colnames(Scaleduscrime)[1] <- "So"  
colnames(Scaleduscrime)[16] <- "Crime"  
  
Scaleduscrime[1:3,]

## So M Ed Po1 Po2 LF M.F  
## 1 1 0.9886930 -1.3085099 -0.9085105 -0.8666988 -1.2667456 -1.1206050  
## 2 0 0.3521372 0.6580587 0.6056737 0.5280852 0.5396568 0.9834175  
## 3 1 0.2725678 -1.4872888 -1.3459415 -1.2958632 -0.6976051 -0.4758239  
## Pop NW U1 U2 Wealth Ineq Prob  
## 1 -0.09500679 1.943738564 0.69510600 0.8313680 -1.3616094 1.679364 1.6497631  
## 2 -0.62033844 0.008483424 0.02950365 0.2393332 0.3276683 0.000000 -0.7693365  
## 3 -0.48900552 1.146296747 -0.08143007 -0.1158877 -2.1492481 1.403647 1.5969416  
## Time Crime  
## 1 -0.05599367 791  
## 2 -0.18315796 1635  
## 3 -0.32416470 578

# Split the data into Training and Test Datasets.  
library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

randomrows <- createDataPartition(y=1:nrow(Scaleduscrime),p=0.7, list = FALSE)  
TrainingData = Scaleduscrime[randomrows,]  
TestData = Scaleduscrime[-randomrows,]  
dim(TrainingData)

## [1] 35 16

dim(TestData)

## [1] 12 16

library(olsrr)

##   
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':  
##   
## rivers

## ---------------------------------------------------------------------------

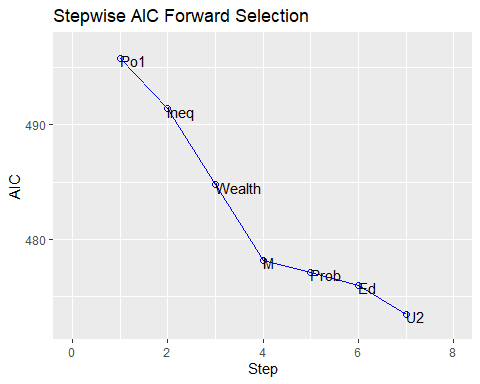
# Perform Forward Regression using aic  
model<-lm(Crime~.,data = TrainingData)  
Forwardfit.aic <-ols\_step\_forward\_aic(model, details = TRUE)

## Forward Selection Method   
## ------------------------  
##   
## Candidate Terms:   
##   
## 1 . So   
## 2 . M   
## 3 . Ed   
## 4 . Po1   
## 5 . Po2   
## 6 . LF   
## 7 . M.F   
## 8 . Pop   
## 9 . NW   
## 10 . U1   
## 11 . U2   
## 12 . Wealth   
## 13 . Ineq   
## 14 . Prob   
## 15 . Time   
##   
## Step 0: AIC = 507.0876   
## Crime ~ 1   
##   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Po1 1 495.841 1128626.869 2453269.016 0.315 0.294   
## Po2 1 496.326 1094409.821 2487486.064 0.306 0.284   
## Prob 1 497.633 999751.467 2582144.419 0.279 0.257   
## Pop 1 499.351 869842.365 2712053.520 0.243 0.220   
## Time 1 502.489 615451.183 2966444.703 0.172 0.147   
## Wealth 1 502.537 611427.125 2970468.761 0.171 0.146   
## U2 1 507.541 154871.969 3427023.917 0.043 0.014   
## Ed 1 507.567 152323.822 3429572.063 0.043 0.014   
## Ineq 1 508.547 54884.610 3527011.276 0.015 -0.015   
## M.F 1 508.844 24841.604 3557054.282 0.007 -0.023   
## So 1 508.940 15112.715 3566783.170 0.004 -0.026   
## NW 1 508.988 10216.386 3571679.500 0.003 -0.027   
## M 1 509.016 7344.408 3574551.478 0.002 -0.028   
## U1 1 509.038 5105.396 3576790.489 0.001 -0.029   
## LF 1 509.085 314.577 3581581.309 0.000 -0.030   
## ------------------------------------------------------------------------------  
##   
##   
## - Po1   
##   
##   
## Step 1 : AIC = 495.8411   
## Crime ~ Po1   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## Ineq 1 491.397 412570.973 2040698.043 0.430 0.395   
## M 1 491.519 405405.508 2047863.508 0.428 0.393   
## Time 1 494.656 213421.255 2239847.762 0.375 0.336   
## So 1 494.849 201012.706 2252256.311 0.371 0.332   
## NW 1 495.057 187592.705 2265676.311 0.367 0.328   
## Prob 1 496.070 121066.144 2332202.872 0.349 0.308   
## Pop 1 496.667 80947.348 2372321.668 0.338 0.296   
## Wealth 1 497.570 18934.023 2434334.994 0.320 0.278   
## M.F 1 497.651 13278.426 2439990.591 0.319 0.276   
## U2 1 497.681 11227.973 2442041.043 0.318 0.276   
## U1 1 497.733 7586.205 2445682.812 0.317 0.275   
## Po2 1 497.746 6676.168 2446592.848 0.317 0.274   
## Ed 1 497.764 5367.968 2447901.049 0.317 0.274   
## LF 1 497.780 4263.764 2449005.252 0.316 0.274   
## -----------------------------------------------------------------------------  
##   
## - Ineq   
##   
##   
## Step 2 : AIC = 491.3966   
## Crime ~ Po1 + Ineq   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## Wealth 1 484.774 445589.256 1595108.787 0.555 0.512   
## Prob 1 486.606 359892.126 1680805.917 0.531 0.485   
## Ed 1 487.472 317758.831 1722939.212 0.519 0.472   
## M.F 1 490.587 157410.744 1883287.299 0.474 0.423   
## M 1 490.768 147630.923 1893067.120 0.471 0.420   
## Time 1 491.324 117321.510 1923376.533 0.463 0.411   
## LF 1 491.955 82341.568 1958356.475 0.453 0.400   
## U1 1 493.363 1980.571 2038717.472 0.431 0.376   
## U2 1 493.379 1023.689 2039674.354 0.431 0.375   
## Pop 1 493.380 951.595 2039746.449 0.431 0.375   
## NW 1 493.390 357.619 2040340.424 0.430 0.375   
## So 1 493.396 41.327 2040656.716 0.430 0.375   
## Po2 1 493.397 2.564 2040695.479 0.430 0.375   
## -----------------------------------------------------------------------------  
##   
## - Wealth   
##   
##   
## Step 3 : AIC = 484.7744   
## Crime ~ Po1 + Ineq + Wealth   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## M 1 478.186 347075.378 1248033.409 0.652 0.605   
## Prob 1 482.112 198933.020 1396175.767 0.610 0.558   
## Time 1 483.484 143109.094 1451999.694 0.595 0.541   
## Ed 1 484.913 82605.957 1512502.830 0.578 0.521   
## M.F 1 485.267 67258.096 1527850.691 0.573 0.517   
## U1 1 486.536 10834.128 1584274.660 0.558 0.499   
## NW 1 486.625 6785.588 1588323.200 0.557 0.497   
## So 1 486.685 4046.061 1591062.727 0.556 0.497   
## LF 1 486.694 3659.550 1591449.238 0.556 0.496   
## U2 1 486.707 3086.595 1592022.192 0.556 0.496   
## Pop 1 486.719 2526.835 1592581.952 0.555 0.496   
## Po2 1 486.745 1337.351 1593771.436 0.555 0.496   
## -----------------------------------------------------------------------------  
##   
## - M   
##   
##   
## Step 4 : AIC = 478.1863   
## Crime ~ Po1 + Ineq + Wealth + M   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## Prob 1 477.124 104553.541 1143479.868 0.681 0.626   
## U1 1 477.869 79970.953 1168062.456 0.674 0.618   
## Ed 1 478.051 73880.461 1174152.948 0.672 0.616   
## U2 1 478.557 56776.806 1191256.603 0.667 0.610   
## M.F 1 478.723 51097.530 1196935.879 0.666 0.608   
## Time 1 478.757 49954.024 1198079.385 0.666 0.608   
## NW 1 479.845 12104.006 1235929.403 0.655 0.595   
## So 1 479.968 7776.684 1240256.725 0.654 0.594   
## LF 1 480.134 1847.025 1246186.383 0.652 0.592   
## Po2 1 480.142 1595.606 1246437.803 0.652 0.592   
## Pop 1 480.144 1494.654 1246538.755 0.652 0.592   
## -----------------------------------------------------------------------------  
##   
## - Prob   
##   
##   
## Step 5 : AIC = 477.124   
## Crime ~ Po1 + Ineq + Wealth + M + Prob   
##   
## ----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ----------------------------------------------------------------------------  
## Ed 1 476.040 96447.490 1047032.378 0.708 0.645   
## U1 1 476.409 85335.284 1058144.584 0.705 0.641   
## U2 1 477.242 59865.135 1083614.733 0.697 0.633   
## M.F 1 477.574 49526.924 1093952.944 0.695 0.629   
## LF 1 478.890 7619.524 1135860.344 0.683 0.615   
## Time 1 478.950 5669.305 1137810.563 0.682 0.614   
## Po2 1 479.035 2913.147 1140566.721 0.682 0.613   
## Pop 1 479.098 841.502 1142638.366 0.681 0.613   
## NW 1 479.105 626.205 1142853.662 0.681 0.613   
## So 1 479.124 7.427 1143472.441 0.681 0.612   
## ----------------------------------------------------------------------------  
##   
## - Ed   
##   
##   
## Step 6 : AIC = 476.04   
## Crime ~ Po1 + Ineq + Wealth + M + Prob + Ed   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## U2 1 473.509 127134.753 919897.625 0.743 0.677   
## U1 1 475.456 74522.160 972510.218 0.728 0.658   
## LF 1 476.002 59212.052 987820.326 0.724 0.653   
## Time 1 477.007 30442.920 1016589.459 0.716 0.643   
## NW 1 477.574 13847.376 1033185.003 0.712 0.637   
## So 1 477.702 10061.297 1036971.081 0.710 0.635   
## M.F 1 477.730 9223.807 1037808.571 0.710 0.635   
## Pop 1 477.942 2924.112 1044108.266 0.709 0.633   
## Po2 1 478.020 596.911 1046435.467 0.708 0.632   
## -----------------------------------------------------------------------------  
##   
## - U2   
##   
##   
## Step 7 : AIC = 473.5091   
## Crime ~ Po1 + Ineq + Wealth + M + Prob + Ed + U2   
##   
## ---------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ---------------------------------------------------------------------------  
## Time 1 474.294 31378.195 888519.430 0.752 0.676   
## LF 1 474.787 18781.605 901116.020 0.748 0.671   
## So 1 475.244 6935.270 912962.355 0.745 0.667   
## NW 1 475.325 4822.279 915075.346 0.745 0.666   
## U1 1 475.449 1579.265 918318.360 0.744 0.665   
## Pop 1 475.492 462.964 919434.661 0.743 0.664   
## M.F 1 475.505 99.901 919797.724 0.743 0.664   
## Po2 1 475.507 50.328 919847.297 0.743 0.664   
## ---------------------------------------------------------------------------  
##   
##   
## No more variables to be added.  
##   
## Variables Entered:   
##   
## - Po1   
## - Ineq   
## - Wealth   
## - M   
## - Prob   
## - Ed   
## - U2   
##   
##   
## Final Model Output   
## ------------------  
##   
## Model Summary   
## -----------------------------------------------------------------  
## R 0.862 RMSE 184.581   
## R-Squared 0.743 Coef. Var 20.622   
## Adj. R-Squared 0.677 MSE 34070.282   
## Pred R-Squared 0.504 MAE 127.287   
## -----------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## -----------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## -----------------------------------------------------------------------  
## Regression 2661998.261 7 380285.466 11.162 0.0000   
## Residual 919897.625 27 34070.282   
## Total 3581895.886 34   
## -----------------------------------------------------------------------  
##   
## Parameter Estimates   
## ---------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ---------------------------------------------------------------------------------------------  
## (Intercept) 882.916 32.809 26.911 0.000 815.598 950.234   
## Po1 164.745 67.323 0.481 2.447 0.021 26.610 302.880   
## Ineq 378.075 75.204 1.183 5.027 0.000 223.769 532.381   
## Wealth 266.440 114.685 0.827 2.323 0.028 31.125 501.755   
## M 140.702 45.849 0.457 3.069 0.005 46.628 234.776   
## Prob -122.299 60.001 -0.321 -2.038 0.051 -245.410 0.813   
## Ed 129.626 59.133 0.408 2.192 0.037 8.294 250.958   
## U2 72.373 37.465 0.217 1.932 0.064 -4.500 149.245   
## ---------------------------------------------------------------------------------------------

Forwardfit.aic

##   
## Selection Summary   
## --------------------------------------------------------------------------  
## Variable AIC Sum Sq RSS R-Sq Adj. R-Sq   
## --------------------------------------------------------------------------  
## Po1 495.841 1128626.869 2453269.016 0.31509 0.29434   
## Ineq 491.397 1541197.843 2040698.043 0.43027 0.39467   
## Wealth 484.774 1986787.099 1595108.787 0.55467 0.51158   
## M 478.186 2333862.477 1248033.409 0.65157 0.60511   
## Prob 477.124 2438416.018 1143479.868 0.68076 0.62572   
## Ed 476.040 2534863.508 1047032.378 0.70769 0.64505   
## U2 473.509 2661998.261 919897.625 0.74318 0.67660   
## --------------------------------------------------------------------------

plot(Forwardfit.aic)



**Analysis : The Forward Selection Model started with no predictors and at each step added one predictor ( selection was based on the least AIC) until the model can no longer be improved.**

**Here the best Model that the Forward Selection Model gave us is Crime ~ M + Ed + Po1 + U2 + Wealth + Ineq + Prob with AIC = 473.509.**

**# Backward Elimination model using aic**  
BackwardFit.aic <- ols\_step\_backward\_aic(model, details = TRUE)

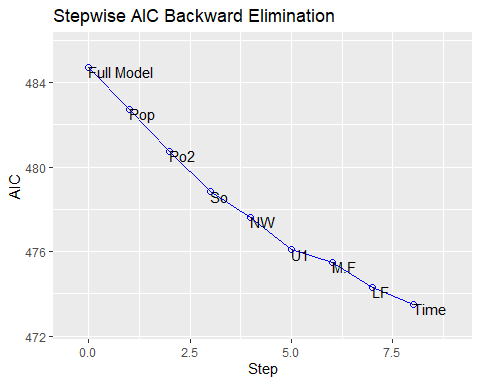
## Backward Elimination Method   
## ---------------------------  
##   
## Candidate Terms:   
##   
## 1 . So   
## 2 . M   
## 3 . Ed   
## 4 . Po1   
## 5 . Po2   
## 6 . LF   
## 7 . M.F   
## 8 . Pop   
## 9 . NW   
## 10 . U1   
## 11 . U2   
## 12 . Wealth   
## 13 . Ineq   
## 14 . Prob   
## 15 . Time   
##   
## Step 0: AIC = 484.7026   
## Crime ~ So + M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 + Wealth + Ineq + Prob + Time   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## Pop 1 482.714 250.498 802109.470 0.776 0.619   
## Po2 1 482.759 1293.475 803152.447 0.776 0.619   
## So 1 482.772 1583.973 803442.946 0.776 0.619   
## Po1 1 483.175 10902.572 812761.545 0.773 0.614   
## NW 1 483.487 18171.034 820030.007 0.771 0.611   
## U1 1 483.552 19705.007 821563.980 0.771 0.610   
## Time 1 484.499 42220.203 844079.176 0.764 0.599   
## U2 1 484.625 45268.764 847127.737 0.763 0.598   
## LF 1 484.721 47608.402 849467.375 0.763 0.597   
## Prob 1 484.765 48676.878 850535.851 0.763 0.596   
## M.F 1 485.236 60195.359 862054.332 0.759 0.591   
## M 1 486.290 86547.099 888406.072 0.752 0.578   
## Wealth 1 486.643 95547.301 897406.274 0.749 0.574   
## Ed 1 489.051 159477.034 961336.007 0.732 0.544   
## Ineq 1 498.583 460403.892 1262262.865 0.648 0.401   
## -----------------------------------------------------------------------------  
##   
##   
## Variables Removed:   
##   
## - Pop   
##   
##   
## Step 1 : AIC = 482.7135   
## Crime ~ So + M + Ed + Po1 + Po2 + LF + M.F + NW + U1 + U2 + Wealth + Ineq + Prob + Time   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## Po2 1 480.761 1092.172 803201.643 0.776 0.637   
## So 1 480.795 1857.929 803967.400 0.776 0.637   
## Po1 1 481.208 11404.397 813513.867 0.773 0.632   
## NW 1 481.517 18630.550 820740.020 0.771 0.629   
## U1 1 481.632 21336.603 823446.074 0.770 0.628   
## U2 1 482.671 46130.102 848239.573 0.763 0.617   
## Time 1 482.682 46415.770 848525.240 0.763 0.616   
## LF 1 482.918 52137.655 854247.125 0.762 0.614   
## Prob 1 483.000 54156.417 856265.887 0.761 0.613   
## M.F 1 483.382 63543.681 865653.151 0.758 0.609   
## M 1 484.493 91466.915 893576.385 0.751 0.596   
## Wealth 1 484.643 95304.099 897413.570 0.749 0.594   
## Ed 1 487.051 159226.701 961336.172 0.732 0.565   
## Ineq 1 497.712 501540.763 1303650.234 0.636 0.411   
## -----------------------------------------------------------------------------  
##   
## - Po2   
##   
##   
## Step 2 : AIC = 480.7612   
## Crime ~ So + M + Ed + Po1 + LF + M.F + NW + U1 + U2 + Wealth + Ineq + Prob + Time   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## So 1 478.826 1489.580 804691.223 0.775 0.653   
## NW 1 479.526 17745.888 820947.531 0.771 0.646   
## U1 1 479.641 20448.211 823649.854 0.770 0.645   
## U2 1 480.675 45147.670 848349.313 0.763 0.634   
## LF 1 480.962 52120.699 855322.342 0.761 0.631   
## Prob 1 481.015 53419.316 856620.959 0.761 0.630   
## Time 1 481.166 57138.700 860340.343 0.760 0.629   
## M.F 1 481.409 63130.298 866331.940 0.758 0.626   
## M 1 482.616 93524.644 896726.287 0.750 0.613   
## Wealth 1 482.900 100836.714 904038.357 0.748 0.610   
## Po1 1 484.060 131278.377 934480.019 0.739 0.597   
## Ed 1 485.551 171951.057 975152.700 0.728 0.579   
## Ineq 1 496.058 513397.425 1316599.068 0.632 0.432   
## -----------------------------------------------------------------------------  
##   
## - So   
##   
##   
## Step 3 : AIC = 478.826   
## Crime ~ M + Ed + Po1 + LF + M.F + NW + U1 + U2 + Wealth + Ineq + Prob + Time   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## NW 1 477.619 18436.438 823127.661 0.770 0.660   
## U1 1 477.683 19940.065 824631.288 0.770 0.660   
## U2 1 478.721 44775.544 849466.766 0.763 0.649   
## Prob 1 479.053 52859.310 857550.533 0.761 0.646   
## LF 1 479.394 61251.778 865943.000 0.758 0.643   
## Time 1 479.412 61702.685 866393.907 0.758 0.642   
## M.F 1 479.425 62022.316 866713.538 0.758 0.642   
## M 1 480.623 92197.866 896889.088 0.750 0.630   
## Wealth 1 480.985 101529.518 906220.741 0.747 0.626   
## Po1 1 482.791 149525.604 954216.827 0.734 0.606   
## Ed 1 483.714 175023.558 979714.780 0.726 0.596   
## Ineq 1 494.433 526093.088 1330784.311 0.628 0.451   
## -----------------------------------------------------------------------------  
##   
## - NW   
##   
##   
## Step 4 : AIC = 477.6188   
## Crime ~ M + Ed + Po1 + LF + M.F + U1 + U2 + Wealth + Ineq + Prob + Time   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## U1 1 476.087 11090.766 834218.426 0.767 0.670   
## Prob 1 477.222 38573.581 861701.242 0.759 0.659   
## U2 1 477.261 39549.547 862677.208 0.759 0.659   
## M.F 1 477.473 44785.628 867913.289 0.758 0.657   
## LF 1 477.653 49263.265 872390.925 0.756 0.655   
## Time 1 478.203 63076.242 886203.902 0.753 0.649   
## Wealth 1 481.272 144278.457 967406.117 0.730 0.617   
## Ed 1 481.717 156678.104 979805.765 0.726 0.612   
## Po1 1 482.173 169529.771 992657.432 0.723 0.607   
## M 1 482.893 190144.934 1013272.594 0.717 0.599   
## Ineq 1 498.234 747540.956 1570668.617 0.561 0.379   
## -----------------------------------------------------------------------------  
##   
## - U1   
##   
##   
## Step 5 : AIC = 476.0873   
## Crime ~ M + Ed + Po1 + LF + M.F + U2 + Wealth + Ineq + Prob + Time   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## M.F 1 475.513 34675.622 868894.049 0.757 0.670   
## Prob 1 475.603 36928.143 871146.569 0.757 0.669   
## U2 1 475.700 39342.180 873560.606 0.756 0.668   
## LF 1 475.733 40167.458 874385.885 0.756 0.668   
## Time 1 476.664 63728.189 897946.616 0.749 0.659   
## Ed 1 479.718 145606.358 979824.784 0.726 0.628   
## Po1 1 480.497 167650.813 1001869.239 0.720 0.620   
## Wealth 1 481.909 208893.018 1043111.445 0.709 0.604   
## M 1 482.058 213358.113 1047576.540 0.708 0.602   
## Ineq 1 499.484 889266.996 1723485.423 0.519 0.346   
## -----------------------------------------------------------------------------  
##   
## - M.F   
##   
##   
## Step 6 : AIC = 475.5127   
## Crime ~ M + Ed + Po1 + LF + U2 + Wealth + Ineq + Prob + Time   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## LF 1 474.294 19625.381 888519.430 0.752 0.676   
## Time 1 474.787 32221.971 901116.020 0.748 0.671   
## Prob 1 476.232 70188.887 939082.935 0.738 0.657   
## U2 1 476.848 86876.467 955770.516 0.733 0.651   
## Po1 1 478.776 141000.520 1009894.568 0.718 0.631   
## Wealth 1 480.625 195800.679 1064694.728 0.703 0.611   
## Ed 1 481.134 211377.795 1080271.844 0.698 0.606   
## M 1 483.066 272702.069 1141596.117 0.681 0.583   
## Ineq 1 497.983 879375.429 1748269.477 0.512 0.362   
## -----------------------------------------------------------------------------  
##   
## - LF   
##   
##   
## Step 7 : AIC = 474.2944   
## Crime ~ M + Ed + Po1 + U2 + Wealth + Ineq + Prob + Time   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## Time 1 473.509 31378.195 919897.625 0.743 0.677   
## Prob 1 474.594 60326.174 948845.604 0.735 0.666   
## U2 1 477.007 128070.028 1016589.459 0.716 0.643   
## Wealth 1 478.818 182046.544 1070565.974 0.701 0.624   
## Po1 1 478.957 186295.061 1074814.491 0.700 0.622   
## Ed 1 479.144 192054.947 1080574.377 0.698 0.620   
## M 1 481.507 267525.139 1156044.570 0.677 0.594   
## Ineq 1 496.044 862794.994 1751314.425 0.511 0.384   
## -----------------------------------------------------------------------------  
##   
## - Time   
##   
##   
## Step 8 : AIC = 473.5091   
## Crime ~ M + Ed + Po1 + U2 + Wealth + Ineq + Prob   
##   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## U2 1 476.040 127134.753 1047032.378 0.708 0.645   
## Prob 1 476.519 141548.376 1061446.001 0.704 0.640   
## Ed 1 477.242 163717.108 1083614.733 0.697 0.633   
## Wealth 1 477.888 183890.591 1103788.216 0.692 0.626   
## Po1 1 478.520 204022.573 1123920.198 0.686 0.619   
## M 1 481.982 320862.878 1240760.503 0.654 0.579   
## Ineq 1 494.632 861088.259 1780985.884 0.503 0.396   
## -----------------------------------------------------------------------------  
##   
##   
## No more variables to be removed.  
##   
## Variables Removed:   
##   
## - Pop   
## - Po2   
## - So   
## - NW   
## - U1   
## - M.F   
## - LF   
## - Time   
##   
##   
## Final Model Output   
## ------------------  
##   
## Model Summary   
## -----------------------------------------------------------------  
## R 0.862 RMSE 184.581   
## R-Squared 0.743 Coef. Var 20.622   
## Adj. R-Squared 0.677 MSE 34070.282   
## Pred R-Squared 0.504 MAE 127.287   
## -----------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## -----------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## -----------------------------------------------------------------------  
## Regression 2661998.261 7 380285.466 11.162 0.0000   
## Residual 919897.625 27 34070.282   
## Total 3581895.886 34   
## -----------------------------------------------------------------------  
##   
## Parameter Estimates   
## ---------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ---------------------------------------------------------------------------------------------  
## (Intercept) 882.916 32.809 26.911 0.000 815.598 950.234   
## M 140.702 45.849 0.457 3.069 0.005 46.628 234.776   
## Ed 129.626 59.133 0.408 2.192 0.037 8.294 250.958   
## Po1 164.745 67.323 0.481 2.447 0.021 26.610 302.880   
## U2 72.373 37.465 0.217 1.932 0.064 -4.500 149.245   
## Wealth 266.440 114.685 0.827 2.323 0.028 31.125 501.755   
## Ineq 378.075 75.204 1.183 5.027 0.000 223.769 532.381   
## Prob -122.299 60.001 -0.321 -2.038 0.051 -245.410 0.813   
## ---------------------------------------------------------------------------------------------

BackwardFit.aic

##   
##   
## Backward Elimination Summary   
## --------------------------------------------------------------------------  
## Variable AIC RSS Sum Sq R-Sq Adj. R-Sq   
## --------------------------------------------------------------------------  
## Full Model 484.703 801858.973 2780036.913 0.77614 0.59940   
## Pop 482.714 802109.470 2779786.415 0.77607 0.61931   
## Po2 480.761 803201.643 2778694.243 0.77576 0.63695   
## So 478.826 804691.223 2777204.663 0.77534 0.65281   
## NW 477.619 823127.661 2758768.225 0.77020 0.66029   
## U1 476.087 834218.426 2747677.459 0.76710 0.67006   
## M.F 475.513 868894.049 2713001.837 0.75742 0.67009   
## LF 474.294 888519.430 2693376.455 0.75194 0.67562   
## Time 473.509 919897.625 2661998.261 0.74318 0.67660   
## --------------------------------------------------------------------------

**Analysis:** The Backward Elimination Summary method removed 8 predictors which are listed above. It started from model with full predictors, at each step it removed the predictor which resulted the highest AIC. It continued until it removed Pop, Po2, So, NW, U1, MF, LF, Time. This is shown below in the Elimination Plot.

plot(BackwardFit.aic)



**Analysis: The Backward Elimination Model, suggests that the Best Model is Crime ~ M + Ed + Po1 + U2 + Wealth + Ineq + Prob with AIC = 473.5091.**

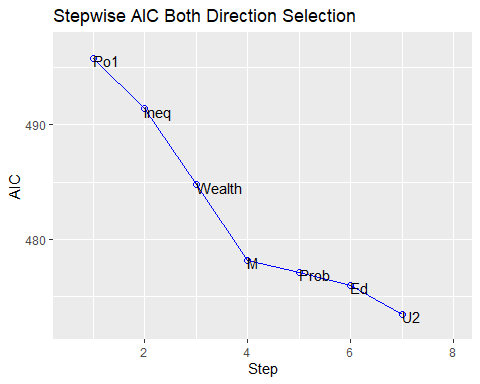
**Stepwise Regression using both directions and aic**   
  
model = model<-lm(Crime~.,data = TrainingData)  
  
StepwiseBothFit.aic<- ols\_step\_both\_aic(model, details = TRUE)

## Stepwise Selection Method   
## -------------------------  
##   
## Candidate Terms:   
##   
## 1 . So   
## 2 . M   
## 3 . Ed   
## 4 . Po1   
## 5 . Po2   
## 6 . LF   
## 7 . M.F   
## 8 . Pop   
## 9 . NW   
## 10 . U1   
## 11 . U2   
## 12 . Wealth   
## 13 . Ineq   
## 14 . Prob   
## 15 . Time   
##   
## Step 0: AIC = 507.0876   
## Crime ~ 1   
##   
##   
## Variables Entered/Removed:   
##   
## Enter New Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Po1 1 495.841 1128626.869 2453269.016 0.315 0.294   
## Po2 1 496.326 1094409.821 2487486.064 0.306 0.284   
## Prob 1 497.633 999751.467 2582144.419 0.279 0.257   
## Pop 1 499.351 869842.365 2712053.520 0.243 0.220   
## Time 1 502.489 615451.183 2966444.703 0.172 0.147   
## Wealth 1 502.537 611427.125 2970468.761 0.171 0.146   
## U2 1 507.541 154871.969 3427023.917 0.043 0.014   
## Ed 1 507.567 152323.822 3429572.063 0.043 0.014   
## Ineq 1 508.547 54884.610 3527011.276 0.015 -0.015   
## M.F 1 508.844 24841.604 3557054.282 0.007 -0.023   
## So 1 508.940 15112.715 3566783.170 0.004 -0.026   
## NW 1 508.988 10216.386 3571679.500 0.003 -0.027   
## M 1 509.016 7344.408 3574551.478 0.002 -0.028   
## U1 1 509.038 5105.396 3576790.489 0.001 -0.029   
## LF 1 509.085 314.577 3581581.309 0.000 -0.030   
## ------------------------------------------------------------------------------  
##   
## - Po1 added   
##   
##   
## Step 1 : AIC = 495.8411   
## Crime ~ Po1   
##   
## Enter New Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Ineq 1 491.397 1541197.843 2040698.043 0.430 0.395   
## M 1 491.519 1534032.377 2047863.508 0.428 0.393   
## Time 1 494.656 1342048.124 2239847.762 0.375 0.336   
## So 1 494.849 1329639.575 2252256.311 0.371 0.332   
## NW 1 495.057 1316219.575 2265676.311 0.367 0.328   
## Prob 1 496.070 1249693.014 2332202.872 0.349 0.308   
## Pop 1 496.667 1209574.218 2372321.668 0.338 0.296   
## Wealth 1 497.570 1147560.892 2434334.994 0.320 0.278   
## M.F 1 497.651 1141905.295 2439990.591 0.319 0.276   
## U2 1 497.681 1139854.842 2442041.043 0.318 0.276   
## U1 1 497.733 1136213.074 2445682.812 0.317 0.275   
## Po2 1 497.746 1135303.037 2446592.848 0.317 0.274   
## Ed 1 497.764 1133994.837 2447901.049 0.317 0.274   
## LF 1 497.780 1132890.633 2449005.252 0.316 0.274   
## ------------------------------------------------------------------------------  
##   
## - Ineq added   
##   
##   
## Step 2 : AIC = 491.3966   
## Crime ~ Po1 + Ineq   
##   
## Remove Existing Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Ineq 1 495.841 1128626.869 2453269.016 0.315 0.294   
## Po1 1 508.547 54884.610 3527011.276 0.015 -0.015   
## ------------------------------------------------------------------------------  
##   
## Enter New Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Wealth 1 484.774 1986787.099 1595108.787 0.555 0.512   
## Prob 1 486.606 1901089.969 1680805.917 0.531 0.485   
## Ed 1 487.472 1858956.673 1722939.212 0.519 0.472   
## M.F 1 490.587 1698608.587 1883287.299 0.474 0.423   
## M 1 490.768 1688828.766 1893067.120 0.471 0.420   
## Time 1 491.324 1658519.352 1923376.533 0.463 0.411   
## LF 1 491.955 1623539.410 1958356.475 0.453 0.400   
## U1 1 493.363 1543178.414 2038717.472 0.431 0.376   
## U2 1 493.379 1542221.531 2039674.354 0.431 0.375   
## Pop 1 493.380 1542149.437 2039746.449 0.431 0.375   
## NW 1 493.390 1541555.461 2040340.424 0.430 0.375   
## So 1 493.396 1541239.170 2040656.716 0.430 0.375   
## Po2 1 493.397 1541200.407 2040695.479 0.430 0.375   
## ------------------------------------------------------------------------------  
##   
## - Wealth added   
##   
##   
## Step 3 : AIC = 484.7744   
## Crime ~ Po1 + Ineq + Wealth   
##   
## Remove Existing Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Po1 1 487.950 1732586.652 1849309.234 0.484 0.451   
## Wealth 1 491.397 1541197.843 2040698.043 0.430 0.395   
## Ineq 1 497.570 1147560.892 2434334.994 0.320 0.278   
## ------------------------------------------------------------------------------  
##   
## Enter New Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## M 1 478.186 2333862.477 1248033.409 0.652 0.605   
## Prob 1 482.112 2185720.119 1396175.767 0.610 0.558   
## Time 1 483.484 2129896.192 1451999.694 0.595 0.541   
## Ed 1 484.913 2069393.056 1512502.830 0.578 0.521   
## M.F 1 485.267 2054045.195 1527850.691 0.573 0.517   
## U1 1 486.536 1997621.226 1584274.660 0.558 0.499   
## NW 1 486.625 1993572.686 1588323.200 0.557 0.497   
## So 1 486.685 1990833.159 1591062.727 0.556 0.497   
## LF 1 486.694 1990446.648 1591449.238 0.556 0.496   
## U2 1 486.707 1989873.693 1592022.192 0.556 0.496   
## Pop 1 486.719 1989313.933 1592581.952 0.555 0.496   
## Po2 1 486.745 1988124.450 1593771.436 0.555 0.496   
## ------------------------------------------------------------------------------  
##   
## - M added   
##   
##   
## Step 4 : AIC = 478.1863   
## Crime ~ Po1 + Ineq + Wealth + M   
##   
## Remove Existing Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Po1 1 482.752 2076346.586 1505549.300 0.580 0.539   
## M 1 484.774 1986787.099 1595108.787 0.555 0.512   
## Wealth 1 490.768 1688828.766 1893067.120 0.471 0.420   
## Ineq 1 492.661 1583622.645 1998273.240 0.442 0.388   
## ------------------------------------------------------------------------------  
##   
## Enter New Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Prob 1 477.124 2438416.018 1143479.868 0.681 0.626   
## U1 1 477.869 2413833.430 1168062.456 0.674 0.618   
## Ed 1 478.051 2407742.938 1174152.948 0.672 0.616   
## U2 1 478.557 2390639.283 1191256.603 0.667 0.610   
## M.F 1 478.723 2384960.007 1196935.879 0.666 0.608   
## Time 1 478.757 2383816.501 1198079.385 0.666 0.608   
## NW 1 479.845 2345966.483 1235929.403 0.655 0.595   
## So 1 479.968 2341639.161 1240256.725 0.654 0.594   
## LF 1 480.134 2335709.502 1246186.383 0.652 0.592   
## Po2 1 480.142 2335458.083 1246437.803 0.652 0.592   
## Pop 1 480.144 2335357.131 1246538.755 0.652 0.592   
## ------------------------------------------------------------------------------  
##   
## - Prob added   
##   
##   
## Step 5 : AIC = 477.124   
## Crime ~ Po1 + Ineq + Wealth + M + Prob   
##   
## Remove Existing Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Prob 1 478.186 2333862.477 1248033.409 0.652 0.605   
## Po1 1 479.725 2277777.427 1304118.458 0.636 0.587   
## M 1 482.112 2185720.119 1396175.767 0.610 0.558   
## Wealth 1 486.619 1993842.141 1588053.745 0.557 0.498   
## Ineq 1 492.983 1677190.103 1904705.783 0.468 0.397   
## ------------------------------------------------------------------------------  
##   
## Enter New Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Ed 1 476.040 2534863.508 1047032.378 0.708 0.645   
## U1 1 476.409 2523751.302 1058144.584 0.705 0.641   
## U2 1 477.242 2498281.153 1083614.733 0.697 0.633   
## M.F 1 477.574 2487942.942 1093952.944 0.695 0.629   
## LF 1 478.890 2446035.542 1135860.344 0.683 0.615   
## Time 1 478.950 2444085.323 1137810.563 0.682 0.614   
## Po2 1 479.035 2441329.165 1140566.721 0.682 0.613   
## Pop 1 479.098 2439257.520 1142638.366 0.681 0.613   
## NW 1 479.105 2439042.223 1142853.662 0.681 0.613   
## So 1 479.124 2438423.445 1143472.441 0.681 0.612   
## ------------------------------------------------------------------------------  
##   
## - Ed added   
##   
##   
## Step 6 : AIC = 476.04   
## Crime ~ Po1 + Ineq + Wealth + M + Prob + Ed   
##   
## Remove Existing Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## Ed 1 477.124 2438416.018 1143479.868 0.681 0.626   
## Prob 1 478.051 2407742.938 1174152.948 0.672 0.616   
## Wealth 1 480.146 2335313.915 1246581.971 0.652 0.592   
## M 1 481.161 2298631.492 1283264.394 0.642 0.580   
## Po1 1 481.348 2291757.391 1290138.495 0.640 0.578   
## Ineq 1 494.862 1683760.669 1898135.217 0.470 0.379   
## ------------------------------------------------------------------------------  
##   
## Enter New Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## U2 1 473.509 2661998.261 919897.625 0.743 0.677   
## U1 1 475.456 2609385.668 972510.218 0.728 0.658   
## LF 1 476.002 2594075.559 987820.326 0.724 0.653   
## Time 1 477.007 2565306.427 1016589.459 0.716 0.643   
## NW 1 477.574 2548710.883 1033185.003 0.712 0.637   
## So 1 477.702 2544924.805 1036971.081 0.710 0.635   
## M.F 1 477.730 2544087.315 1037808.571 0.710 0.635   
## Pop 1 477.942 2537787.620 1044108.266 0.709 0.633   
## Po2 1 478.020 2535460.419 1046435.467 0.708 0.632   
## ------------------------------------------------------------------------------  
##   
## - U2 added   
##   
##   
## Step 7 : AIC = 473.5091   
## Crime ~ Po1 + Ineq + Wealth + M + Prob + Ed + U2   
##   
## Remove Existing Variables   
## ------------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## ------------------------------------------------------------------------------  
## U2 1 476.040 2534863.508 1047032.378 0.708 0.645   
## Prob 1 476.519 2520449.884 1061446.001 0.704 0.640   
## Ed 1 477.242 2498281.153 1083614.733 0.697 0.633   
## Wealth 1 477.888 2478107.670 1103788.216 0.692 0.626   
## Po1 1 478.520 2457975.688 1123920.198 0.686 0.619   
## M 1 481.982 2341135.382 1240760.503 0.654 0.579   
## Ineq 1 494.632 1800910.001 1780985.884 0.503 0.396   
## ------------------------------------------------------------------------------  
##   
## Enter New Variables   
## -----------------------------------------------------------------------------  
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq   
## -----------------------------------------------------------------------------  
## Time 1 474.294 2693376.455 888519.430 0.752 0.676   
## LF 1 474.787 2680779.866 901116.020 0.748 0.671   
## So 1 475.244 2668933.530 912962.355 0.745 0.667   
## NW 1 475.325 2666820.540 915075.346 0.745 0.666   
## U1 1 475.449 2663577.526 918318.360 0.744 0.665   
## Pop 1 475.492 2662461.225 919434.661 0.743 0.664   
## M.F 1 475.505 2662098.161 919797.724 0.743 0.664   
## Po2 1 475.507 2662048.589 919847.297 0.743 0.664   
## -----------------------------------------------------------------------------  
##   
##   
## No more variables to be added or removed.  
##   
## Final Model Output   
## ------------------  
##   
## Model Summary   
## -----------------------------------------------------------------  
## R 0.862 RMSE 184.581   
## R-Squared 0.743 Coef. Var 20.622   
## Adj. R-Squared 0.677 MSE 34070.282   
## Pred R-Squared 0.504 MAE 127.287   
## -----------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## -----------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## -----------------------------------------------------------------------  
## Regression 2661998.261 7 380285.466 11.162 0.0000   
## Residual 919897.625 27 34070.282   
## Total 3581895.886 34   
## -----------------------------------------------------------------------  
##   
## Parameter Estimates   
## ---------------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ---------------------------------------------------------------------------------------------  
## (Intercept) 882.916 32.809 26.911 0.000 815.598 950.234   
## Po1 164.745 67.323 0.481 2.447 0.021 26.610 302.880   
## Ineq 378.075 75.204 1.183 5.027 0.000 223.769 532.381   
## Wealth 266.440 114.685 0.827 2.323 0.028 31.125 501.755   
## M 140.702 45.849 0.457 3.069 0.005 46.628 234.776   
## Prob -122.299 60.001 -0.321 -2.038 0.051 -245.410 0.813   
## Ed 129.626 59.133 0.408 2.192 0.037 8.294 250.958   
## U2 72.373 37.465 0.217 1.932 0.064 -4.500 149.245   
## ---------------------------------------------------------------------------------------------

StepwiseBothFit.aic

##   
##   
## Stepwise Summary   
## -------------------------------------------------------------------------------------  
## Variable Method AIC RSS Sum Sq R-Sq Adj. R-Sq   
## -------------------------------------------------------------------------------------  
## Po1 addition 495.841 2453269.016 1128626.869 0.31509 0.29434   
## Ineq addition 491.397 2040698.043 1541197.843 0.43027 0.39467   
## Wealth addition 484.774 1595108.787 1986787.099 0.55467 0.51158   
## M addition 478.186 1248033.409 2333862.477 0.65157 0.60511   
## Prob addition 477.124 1143479.868 2438416.018 0.68076 0.62572   
## Ed addition 476.040 1047032.378 2534863.508 0.70769 0.64505   
## U2 addition 473.509 919897.625 2661998.261 0.74318 0.67660   
## -------------------------------------------------------------------------------------

plot(StepwiseBothFit.aic)



**#Analysis:**

**Stepwise Regression is a combination of both Forward and Backward Regression.**

**All the three methods, Forward Regression, Backward Elimination and Stepwise Regression in both Directions returned the model with lm(Crime~Po1 + Ineq + Wealth + Prob + M + Ed + U2 ) using the Scaled Training Data.**  
  
BestModelWithTrainingData<- lm(Crime~Po1 + Ineq + Wealth + M + Ed + U2 +Prob, data = Training Data)  
summary(BestModelWithTrainingData)

##   
## Call:  
## lm(formula = Crime ~ Po1 + Ineq + Wealth + M + Ed + U2 + Prob,   
## data = Training Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -267.58 -125.25 -6.28 96.12 451.18   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 882.92 32.81 26.911 < 2e-16 \*\*\*  
## Po1 164.74 67.32 2.447 0.02119 \*   
## Ineq 378.08 75.20 5.027 2.83e-05 \*\*\*  
## Wealth 266.44 114.69 2.323 0.02794 \*   
## M 140.70 45.85 3.069 0.00485 \*\*   
## Ed 129.63 59.13 2.192 0.03717 \*   
## U2 72.37 37.47 1.932 0.06395 .   
## Prob -122.30 60.00 -2.038 0.05143 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 184.6 on 27 degrees of freedom  
## Multiple R-squared: 0.7432, Adjusted R-squared: 0.6766   
## F-statistic: 11.16 on 7 and 27 DF, p-value: 1.476e-06

BestModelwithTestData<- lm(Crime~Po1 + Ineq + Wealth + M + Ed + U2 + Prob, data = TestData)  
summary(BestModelwithTestData)

##   
## Call:  
## lm(formula = Crime ~ Po1 + Ineq + Wealth + M + Ed + U2 + Prob,   
## data = TestData)  
##   
## Residuals:  
## 2 5 8 18 21 22 25   
## 163.97555 -73.52705 81.19548 57.51035 0.72107 -64.95555 -0.07717   
## 26 27 42 43 45   
## -128.98342 -28.64515 20.95836 15.66399 -43.83645   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 919.86 43.57 21.111 2.98e-05 \*\*\*  
## Po1 399.55 60.51 6.604 0.00273 \*\*   
## Ineq 108.45 102.27 1.060 0.34872   
## Wealth -113.01 100.27 -1.127 0.32277   
## M 131.94 62.07 2.126 0.10070   
## Ed 294.87 94.78 3.111 0.03583 \*   
## U2 163.74 50.80 3.223 0.03218 \*   
## Prob -86.44 35.89 -2.409 0.07366 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 128.9 on 4 degrees of freedom  
## Multiple R-squared: 0.9798, Adjusted R-squared: 0.9444   
## F-statistic: 27.67 on 7 and 4 DF, p-value: 0.003117

**Analysis:**

**The Initial Model had 15 predictors. The Stepwise Selection model, forward regression and backward elimination models , all these three methods removed the same predictors and suggested best model as Crime~Po1 + Ineq + Wealth + Prob + M + Ed + U2 using the Scaled Training Data.**

**The Adjusted R^2 for Training Data Set is 0.6766 While for the test Data Set it is *0.9444. This shows that the Predictor elimination using stepwise Regression has lowered the Variance.***

**b. Lasso Regression**

**c. Elastic.Net Regression**

library(glmnet)

## Loading required package: Matrix

## Loaded glmnet 4.0-2

set.seed(82)  
  
# The cv part means we want to use Cross validation to   
#obtain the optimalvalues for Lambda.  
model\_lasso <- cv.glmnet(x=as.matrix(TrainingData[,-16]),  
 y = as.matrix(TrainingData[,16]),  
 alpha = 1 ,  
 nfolds = 8,  
 nlambda = 20,  
 type.measure = "mse",  
 family ="gaussian",  
 standardize = TRUE)  
model\_lasso.predicted<-predict(model\_lasso,s=model\_lasso$lambda.1se,newx=as.matrix(TestData[,-16]))  
#Lambda.1se is the value of lambda,that resulted in the simplest model(model with few non zero parameters)  
#and was within 1 standard error of the lambda that had the smallest sum.  
model\_lasso.predicted

## 1  
## 2 895.0571  
## 5 895.0571  
## 8 895.0571  
## 18 895.0571  
## 21 895.0571  
## 22 895.0571  
## 25 895.0571  
## 26 895.0571  
## 27 895.0571  
## 42 895.0571  
## 43 895.0571  
## 45 895.0571

# Find the accuracy   
sse = sum((model\_lasso.predicted - TestData[,16])^2)  
totalSumofSquares = sum((TestData[,16]-mean(TestData[,16]))^2)  
RSquared = 1- (sse/totalSumofSquares)  
RSquared

## [1] -0.005634715

#Elastic.Net Regression   
  
# The cv part means we want to use Cross validation to   
#obtain the optimalvalues for Lambda.  
model\_elasticnet\_alpha0.5 <- cv.glmnet(x=as.matrix(TrainingData[,-16]),  
 y = as.matrix(TrainingData[,16]),  
 alpha = 0.5 ,  
 nfolds = 8,  
 nlambda = 20,  
 type.measure = "mse",  
 family ="gaussian",  
 standardize = TRUE)  
model\_elasticnet\_alpha0.5.predicted<-predict(model\_elasticnet\_alpha0.5,s=model\_elasticnet\_alpha0.5$lambda.1se,newx=as.matrix(TestData[,-16]))  
#Lambda.1se is the value of lambda,that resulted in the simplest model(model with few non zero parameters)  
#and was within 1 standard error of the lambda that had the smallest sum.  
model\_elasticnet\_alpha0.5.predicted

## 1  
## 2 1111.1050  
## 5 971.0996  
## 8 1087.6278  
## 18 691.4622  
## 21 928.0997  
## 22 743.8468  
## 25 568.1519  
## 26 1292.4630  
## 27 631.7364  
## 42 357.4658  
## 43 1015.9012  
## 45 772.1347

# Find the accuracy   
sse = sum((model\_elasticnet\_alpha0.5.predicted - TestData[,16])^2)  
totalSumofSquares = sum((TestData[,16]-mean(TestData[,16]))^2)  
RSquared = 1- (sse/totalSumofSquares)  
AdjustedRSqaured = RSquared - (1-RSquared)\*15/(nrow(TestData)-15-1)  
AdjustedRSqaured

## [1] 2.25108

RSquared

## [1] 0.5450617

#Lets try more values of alpha  
# We create the Elastic.NET fit using the cv.glmnet() function,  
#which takes alpha values from 0.0,0.1,..1.0.  
list\_of\_fits <- list()  
for(i in 0:10)  
{  
 fit.name <- paste0("alpha",i/10)  
 list\_of\_fits[[fit.name]] <- cv.glmnet(x=as.matrix(TrainingData[,-16]),  
 y = as.matrix(TrainingData[,16]),  
 alpha = i/10 ,  
 nfolds = 8,  
 nlambda = 20,  
 type.measure = "mse",  
 family ="gaussian",  
 standardize = FALSE)  
}  
  
results <- data.frame()  
# This for loop will give us the error values for each model from above.  
for(i in 0:10)  
{  
 fit.name <- paste0("alpha",i/10)  
 predicted <- predict(list\_of\_fits[[fit.name]],  
 s=list\_of\_fits[[fit.name]]$lambda.1se,newx=as.matrix(TestData[,-16]))  
   
 # Find the accuracy   
 sse = sum((predicted - TestData[,16])^2)  
 totalSumofSquares = sum((TestData[,16]-mean(TestData[,16]))^2)  
 RSquared = 1- (sse/totalSumofSquares)  
 temp <- data.frame(alpha=i/10, Rsqaured=RSquared, fit.name)  
 results <- rbind(results, temp)  
}  
  
results

## alpha Rsqaured fit.name  
## 1 0.0 0.192802377 alpha0  
## 2 0.1 0.151978925 alpha0.1  
## 3 0.2 -0.005634715 alpha0.2  
## 4 0.3 -0.005634715 alpha0.3  
## 5 0.4 -0.005634715 alpha0.4  
## 6 0.5 0.281999903 alpha0.5  
## 7 0.6 0.151110822 alpha0.6  
## 8 0.7 -0.005634715 alpha0.7  
## 9 0.8 -0.005634715 alpha0.8  
## 10 0.9 0.218213133 alpha0.9  
## 11 1.0 -0.005634715 alpha1

model\_elasticnet\_alpha0.5$glmnet.fit

##   
## Call: glmnet(x = as.matrix(TrainingData[, -16]), y = as.matrix(TrainingData[, 16]), alpha = 0.5, nlambda = 20, family = "gaussian", standardize = TRUE)   
##   
## Df %Dev Lambda  
## 1 0 0.00 359.10  
## 2 4 19.67 221.20  
## 3 5 31.22 136.20  
## 4 7 38.91 83.89  
## 5 13 51.28 51.66  
## 6 13 63.84 31.82  
## 7 13 70.37 19.59  
## 8 14 73.82 12.07  
## 9 15 75.76 7.43  
## 10 15 76.79 4.58  
## 11 14 77.25 2.82  
## 12 13 77.42 1.74  
## 13 13 77.51 1.07  
## 14 14 77.55 0.66  
## 15 14 77.57 0.41  
## 16 15 77.58 0.25  
## 17 15 77.60 0.15  
## 18 15 77.61 0.09  
## 19 15 77.61 0.06  
## 20 15 77.61 0.04

**Analysis:**

For Using Lasso and Elastic.Net Regression in R, we used glmnet library.

Lasso Regression Penalty = Sum of Squared Residuals + Lambda1(|var1| + … + |varx|) + Lambda2(var1^2+…+varx^2).

Glmnet interprets these Lambda’s differently.

Glmnet has a single Lambda as shown:

Regression Penalty =Sum of Squared Residuals +

Lambda\*[ alpha\*(|var1| +…+ |varx|) + (1-alpha) (var1^2+..varx^2)].

When alpha = 0, Lasso penalty goes to zero and the model reduces to ridge regression.

When alpha = 1, Rigde regression penalty goes to zero and the model reduces to Lasso regression.

When 0< alpha < 1 , then the model reduces to Elastic.Net.

Lambda controls how much penalty to apply to the regression.

When Lambda = 0, the model reduces to Linear Regression as penalty = 0.

When Lambda > 0, then Elastis.Net penalty kicks in.

Run glmnet for finding Lasso and ElasticNet Best Models and these results are obtained

## alpha Rsqaured fit.name  
## 1 0.0 0.192802377 alpha0 -> This fit is Ridge Regression  
## 2 0.1 0.151978925 alpha0.1 -> This fit is Elastic.Net Regression  
## 3 0.2 -0.005634715 alpha0.2 -> This fit is Elastic.Net Regression   
## 4 0.3 -0.005634715 alpha0.3 -> This fit is Elastic.Net Regression  
## 5 0.4 -0.005634715 alpha0.4 -> This fit is Elastic.Net Regression  
## 6 0.5 0.281999903 alpha0.5 -> This fit is Elastic.Net Regression  
## 7 0.6 0.151110822 alpha0.6 -> This fit is Elastic.Net Regression  
## 8 0.7 -0.005634715 alpha0.7 -> This fit is Elastic.Net Regression  
## 9 0.8 -0.005634715 alpha0.8 -> This fit is Elastic.Net Regression  
## 10 0.9 0.218213133 alpha0.9 -> This fit is Elastic.Net Regression  
## 11 1.0 -0.005634715 alpha1 -> This fit is Lasso Regression

The Highest RSquared error is for model fit with alpha = 0.5 which is an Elastic.Net Regression. This model seems to be better than the rest.

Next best model is model fit with alpha = 0.9 which is also from Elastic.net Regression.

The Alpha = 0 is the Ridge Regression Model and Alpha = 1 is the Lasso Model.

Based on the Rsquared, Elastic.Net wins, then Ridge Regression is the next best and the next will be Lasso.

For the Best Elastic.Net Model, the lambda values ranged between 0.04 to 359.10 in the cv model.