Model Selection for Vehicle MPG

Preston Tolbert

4/6/2020

### 

### Abstract

In this project, I sought out to find the factors that contributed most to mpg in vehicles. The dataset I used was the mtcars data. After running a regression using all variables, I used model selection techniques to determine the best model.

### Materials and Methods

The main goal of the project is to determine the factors that have the largest impact on mpg in vehiceles. The dataset that will be used to determine this is the mtcars dataset. It is a list of 32 vehicles of varying sizes and performance, therefore the dataset should be a fairly representative sample.

data(mtcars)  
summary(mtcars)

## mpg cyl disp hp   
## Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0   
## 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5   
## Median :19.20 Median :6.000 Median :196.3 Median :123.0   
## Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7   
## 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0   
## Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0   
## drat wt qsec vs   
## Min. :2.760 Min. :1.513 Min. :14.50 Min. :0.0000   
## 1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000   
## Median :3.695 Median :3.325 Median :17.71 Median :0.0000   
## Mean :3.597 Mean :3.217 Mean :17.85 Mean :0.4375   
## 3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000   
## Max. :4.930 Max. :5.424 Max. :22.90 Max. :1.0000   
## am gear carb   
## Min. :0.0000 Min. :3.000 Min. :1.000   
## 1st Qu.:0.0000 1st Qu.:3.000 1st Qu.:2.000   
## Median :0.0000 Median :4.000 Median :2.000   
## Mean :0.4062 Mean :3.688 Mean :2.812   
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :1.0000 Max. :5.000 Max. :8.000

I will first run a regression on the dataset to determine what variables contribute to mpg. The initial model will be the full model, including all measured variables in the dataset. This is done in order to allow our model selection processes to have the most amount of information to make the correct decisions.

lmod<-lm(mpg~cyl+disp+hp+drat+wt+qsec+vs+am+gear  
 +carb, data=mtcars)  
summary(lmod)

##   
## Call:  
## lm(formula = mpg ~ cyl + disp + hp + drat + wt + qsec + vs +   
## am + gear + carb, data = mtcars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.4506 -1.6044 -0.1196 1.2193 4.6271   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 12.30337 18.71788 0.657 0.5181   
## cyl -0.11144 1.04502 -0.107 0.9161   
## disp 0.01334 0.01786 0.747 0.4635   
## hp -0.02148 0.02177 -0.987 0.3350   
## drat 0.78711 1.63537 0.481 0.6353   
## wt -3.71530 1.89441 -1.961 0.0633 .  
## qsec 0.82104 0.73084 1.123 0.2739   
## vs 0.31776 2.10451 0.151 0.8814   
## am 2.52023 2.05665 1.225 0.2340   
## gear 0.65541 1.49326 0.439 0.6652   
## carb -0.19942 0.82875 -0.241 0.8122   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.65 on 21 degrees of freedom  
## Multiple R-squared: 0.869, Adjusted R-squared: 0.8066   
## F-statistic: 13.93 on 10 and 21 DF, p-value: 3.793e-07

For the model selection, I will use forward selection, backward selection, and stepwise selection. I use all of these since they use different criteria in order to select the most important variables. Once all selection processes are completed, I will compare them and select the best model.

ols\_step\_forward\_p(lmod,details=TRUE)

## Forward Selection Method   
## ---------------------------  
##   
## Candidate Terms:   
##   
## 1. cyl   
## 2. disp   
## 3. hp   
## 4. drat   
## 5. wt   
## 6. qsec   
## 7. vs   
## 8. am   
## 9. gear   
## 10. carb   
##   
## We are selecting variables based on p value...  
##   
##   
## Forward Selection: Step 1   
##   
## - wt   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.868 RMSE 3.046   
## R-Squared 0.753 Coef. Var 15.161   
## Adj. R-Squared 0.745 MSE 9.277   
## Pred R-Squared 0.709 MAE 2.341   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 847.725 1 847.725 91.375 0.0000   
## Residual 278.322 30 9.277   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------  
## (Intercept) 37.285 1.878 19.858 0.000 33.450 41.120   
## wt -5.344 0.559 -0.868 -9.559 0.000 -6.486 -4.203   
## ----------------------------------------------------------------------------------------  
##   
##   
##   
## Forward Selection: Step 2   
##   
## - cyl   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.911 RMSE 2.568   
## R-Squared 0.830 Coef. Var 12.780   
## Adj. R-Squared 0.819 MSE 6.592   
## Pred R-Squared 0.790 MAE 1.921   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 934.875 2 467.438 70.908 0.0000   
## Residual 191.172 29 6.592   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------  
## (Intercept) 39.686 1.715 23.141 0.000 36.179 43.194   
## wt -3.191 0.757 -0.518 -4.216 0.000 -4.739 -1.643   
## cyl -1.508 0.415 -0.447 -3.636 0.001 -2.356 -0.660   
## ----------------------------------------------------------------------------------------  
##   
##   
##   
## Forward Selection: Step 3   
##   
## - hp   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.918 RMSE 2.512   
## R-Squared 0.843 Coef. Var 12.501   
## Adj. R-Squared 0.826 MSE 6.308   
## Pred R-Squared 0.796 MAE 1.845   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 949.427 3 316.476 50.171 0.0000   
## Residual 176.621 28 6.308   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------  
## (Intercept) 38.752 1.787 21.687 0.000 35.092 42.412   
## wt -3.167 0.741 -0.514 -4.276 0.000 -4.684 -1.650   
## cyl -0.942 0.551 -0.279 -1.709 0.098 -2.070 0.187   
## hp -0.018 0.012 -0.205 -1.519 0.140 -0.042 0.006   
## ----------------------------------------------------------------------------------------  
##   
##   
##   
## No more variables to be added.  
##   
## Variables Entered:   
##   
## + wt   
## + cyl   
## + hp   
##   
##   
## Final Model Output   
## ------------------  
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.918 RMSE 2.512   
## R-Squared 0.843 Coef. Var 12.501   
## Adj. R-Squared 0.826 MSE 6.308   
## Pred R-Squared 0.796 MAE 1.845   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 949.427 3 316.476 50.171 0.0000   
## Residual 176.621 28 6.308   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------  
## (Intercept) 38.752 1.787 21.687 0.000 35.092 42.412   
## wt -3.167 0.741 -0.514 -4.276 0.000 -4.684 -1.650   
## cyl -0.942 0.551 -0.279 -1.709 0.098 -2.070 0.187   
## hp -0.018 0.012 -0.205 -1.519 0.140 -0.042 0.006   
## ----------------------------------------------------------------------------------------

##   
## Selection Summary   
## -------------------------------------------------------------------------  
## Variable Adj.   
## Step Entered R-Square R-Square C(p) AIC RMSE   
## -------------------------------------------------------------------------  
## 1 wt 0.7528 0.7446 11.6270 166.0294 3.0459   
## 2 cyl 0.8302 0.8185 1.2187 156.0101 2.5675   
## 3 hp 0.8431 0.8263 1.1469 155.4766 2.5115   
## -------------------------------------------------------------------------

ols\_step\_backward\_p(lmod,details=TRUE)

## Backward Elimination Method   
## ---------------------------  
##   
## Candidate Terms:   
##   
## 1 . cyl   
## 2 . disp   
## 3 . hp   
## 4 . drat   
## 5 . wt   
## 6 . qsec   
## 7 . vs   
## 8 . am   
## 9 . gear   
## 10 . carb   
##   
## We are eliminating variables based on p value...  
##   
## - cyl   
##   
## Backward Elimination: Step 1   
##   
## Variable cyl Removed   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.932 RMSE 2.590   
## R-Squared 0.869 Coef. Var 12.891   
## Adj. R-Squared 0.815 MSE 6.708   
## Pred R-Squared 0.704 MAE 1.720   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 978.473 9 108.719 16.208 0.0000   
## Residual 147.574 22 6.708   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## -----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## -----------------------------------------------------------------------------------------  
## (Intercept) 10.960 13.530 0.810 0.427 -17.100 39.020   
## disp 0.013 0.017 0.264 0.763 0.454 -0.022 0.048   
## hp -0.022 0.021 -0.249 -1.048 0.306 -0.065 0.021   
## drat 0.835 1.536 0.074 0.544 0.592 -2.351 4.021   
## wt -3.693 1.840 -0.599 -2.007 0.057 -7.507 0.122   
## qsec 0.842 0.687 0.250 1.227 0.233 -0.582 2.267   
## vs 0.390 1.948 0.033 0.200 0.843 -3.650 4.430   
## am 2.577 1.940 0.213 1.328 0.198 -1.447 6.601   
## gear 0.712 1.366 0.087 0.521 0.608 -2.121 3.544   
## carb -0.220 0.789 -0.059 -0.278 0.783 -1.855 1.416   
## -----------------------------------------------------------------------------------------  
##   
##   
## - vs   
##   
## Backward Elimination: Step 2   
##   
## Variable vs Removed   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.932 RMSE 2.535   
## R-Squared 0.869 Coef. Var 12.620   
## Adj. R-Squared 0.823 MSE 6.428   
## Pred R-Squared 0.732 MAE 1.741   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 978.204 8 122.276 19.022 0.0000   
## Residual 147.843 23 6.428   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## -----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## -----------------------------------------------------------------------------------------  
## (Intercept) 9.768 11.892 0.821 0.420 -14.833 34.369   
## disp 0.012 0.016 0.250 0.753 0.459 -0.021 0.045   
## hp -0.021 0.020 -0.238 -1.051 0.304 -0.062 0.020   
## drat 0.875 1.491 0.078 0.587 0.563 -2.210 3.960   
## wt -3.712 1.798 -0.603 -2.064 0.050 -7.432 0.009   
## qsec 0.911 0.583 0.270 1.562 0.132 -0.295 2.117   
## am 2.524 1.881 0.209 1.342 0.193 -1.368 6.416   
## gear 0.760 1.316 0.093 0.577 0.569 -1.962 3.482   
## carb -0.248 0.759 -0.066 -0.327 0.747 -1.819 1.323   
## -----------------------------------------------------------------------------------------  
##   
##   
## - carb   
##   
## Backward Elimination: Step 3   
##   
## Variable carb Removed   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.932 RMSE 2.488   
## R-Squared 0.868 Coef. Var 12.382   
## Adj. R-Squared 0.830 MSE 6.189   
## Pred R-Squared 0.762 MAE 1.743   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 977.519 7 139.646 22.565 0.0000   
## Residual 148.528 24 6.189   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## -----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## -----------------------------------------------------------------------------------------  
## (Intercept) 9.198 11.542 0.797 0.433 -14.624 33.020   
## disp 0.016 0.012 0.319 1.278 0.213 -0.010 0.041   
## hp -0.025 0.016 -0.281 -1.548 0.135 -0.058 0.008   
## drat 0.810 1.450 0.072 0.559 0.582 -2.183 3.803   
## wt -4.131 1.236 -0.671 -3.342 0.003 -6.681 -1.580   
## qsec 1.010 0.489 0.299 2.066 0.050 0.001 2.019   
## am 2.590 1.835 0.214 1.411 0.171 -1.198 6.378   
## gear 0.606 1.206 0.074 0.503 0.620 -1.883 3.095   
## -----------------------------------------------------------------------------------------  
##   
##   
## - gear   
##   
## Backward Elimination: Step 4   
##   
## Variable gear Removed   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.931 RMSE 2.450   
## R-Squared 0.867 Coef. Var 12.196   
## Adj. R-Squared 0.835 MSE 6.004   
## Pred R-Squared 0.785 MAE 1.769   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 975.954 6 162.659 27.093 0.0000   
## Residual 150.093 25 6.004   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## -----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## -----------------------------------------------------------------------------------------  
## (Intercept) 10.711 10.975 0.976 0.338 -11.894 33.315   
## disp 0.013 0.011 0.269 1.193 0.244 -0.010 0.036   
## hp -0.022 0.015 -0.248 -1.488 0.149 -0.052 0.008   
## drat 1.021 1.367 0.091 0.746 0.462 -1.796 3.837   
## wt -4.045 1.206 -0.657 -3.355 0.003 -6.527 -1.562   
## qsec 0.991 0.480 0.294 2.064 0.050 0.002 1.979   
## am 2.985 1.634 0.247 1.827 0.080 -0.380 6.350   
## -----------------------------------------------------------------------------------------  
##   
##   
## - drat   
##   
## Backward Elimination: Step 5   
##   
## Variable drat Removed   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.929 RMSE 2.429   
## R-Squared 0.864 Coef. Var 12.092   
## Adj. R-Squared 0.838 MSE 5.901   
## Pred R-Squared 0.798 MAE 1.815   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 972.609 5 194.522 32.962 0.0000   
## Residual 153.438 26 5.901   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------  
## (Intercept) 14.362 9.741 1.474 0.152 -5.661 34.384   
## disp 0.011 0.011 0.231 1.060 0.299 -0.011 0.033   
## hp -0.021 0.015 -0.241 -1.460 0.156 -0.051 0.009   
## wt -4.084 1.194 -0.663 -3.420 0.002 -6.539 -1.630   
## qsec 1.007 0.475 0.299 2.118 0.044 0.030 1.984   
## am 3.470 1.486 0.287 2.336 0.027 0.416 6.525   
## ----------------------------------------------------------------------------------------  
##   
##   
##   
## No more variables satisfy the condition of p value = 0.3  
##   
##   
## Variables Removed:   
##   
## - cyl   
## - vs   
## - carb   
## - gear   
## - drat   
##   
##   
## Final Model Output   
## ------------------  
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.929 RMSE 2.429   
## R-Squared 0.864 Coef. Var 12.092   
## Adj. R-Squared 0.838 MSE 5.901   
## Pred R-Squared 0.798 MAE 1.815   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 972.609 5 194.522 32.962 0.0000   
## Residual 153.438 26 5.901   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------  
## (Intercept) 14.362 9.741 1.474 0.152 -5.661 34.384   
## disp 0.011 0.011 0.231 1.060 0.299 -0.011 0.033   
## hp -0.021 0.015 -0.241 -1.460 0.156 -0.051 0.009   
## wt -4.084 1.194 -0.663 -3.420 0.002 -6.539 -1.630   
## qsec 1.007 0.475 0.299 2.118 0.044 0.030 1.984   
## am 3.470 1.486 0.287 2.336 0.027 0.416 6.525   
## ----------------------------------------------------------------------------------------

##   
##   
## Elimination Summary   
## ------------------------------------------------------------------------  
## Variable Adj.   
## Step Removed R-Square R-Square C(p) AIC RMSE   
## ------------------------------------------------------------------------  
## 1 cyl 0.8689 0.8153 9.0114 161.7271 2.5900   
## 2 vs 0.8687 0.823 7.0496 159.7853 2.5353   
## 3 carb 0.8681 0.8296 5.1472 157.9333 2.4877   
## 4 gear 0.8667 0.8347 3.3700 156.2687 2.4503   
## 5 drat 0.8637 0.8375 1.8462 154.9740 2.4293   
## ------------------------------------------------------------------------

ols\_step\_both\_p(lmod,details=TRUE)

## Stepwise Selection Method   
## ---------------------------  
##   
## Candidate Terms:   
##   
## 1. cyl   
## 2. disp   
## 3. hp   
## 4. drat   
## 5. wt   
## 6. qsec   
## 7. vs   
## 8. am   
## 9. gear   
## 10. carb   
##   
## We are selecting variables based on p value...  
##   
##   
## Stepwise Selection: Step 1   
##   
## - wt added   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.868 RMSE 3.046   
## R-Squared 0.753 Coef. Var 15.161   
## Adj. R-Squared 0.745 MSE 9.277   
## Pred R-Squared 0.709 MAE 2.341   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 847.725 1 847.725 91.375 0.0000   
## Residual 278.322 30 9.277   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------  
## (Intercept) 37.285 1.878 19.858 0.000 33.450 41.120   
## wt -5.344 0.559 -0.868 -9.559 0.000 -6.486 -4.203   
## ----------------------------------------------------------------------------------------  
##   
##   
##   
## Stepwise Selection: Step 2   
##   
## - cyl added   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.911 RMSE 2.568   
## R-Squared 0.830 Coef. Var 12.780   
## Adj. R-Squared 0.819 MSE 6.592   
## Pred R-Squared 0.790 MAE 1.921   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 934.875 2 467.438 70.908 0.0000   
## Residual 191.172 29 6.592   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------  
## (Intercept) 39.686 1.715 23.141 0.000 36.179 43.194   
## wt -3.191 0.757 -0.518 -4.216 0.000 -4.739 -1.643   
## cyl -1.508 0.415 -0.447 -3.636 0.001 -2.356 -0.660   
## ----------------------------------------------------------------------------------------  
##   
##   
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.911 RMSE 2.568   
## R-Squared 0.830 Coef. Var 12.780   
## Adj. R-Squared 0.819 MSE 6.592   
## Pred R-Squared 0.790 MAE 1.921   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 934.875 2 467.438 70.908 0.0000   
## Residual 191.172 29 6.592   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------  
## (Intercept) 39.686 1.715 23.141 0.000 36.179 43.194   
## wt -3.191 0.757 -0.518 -4.216 0.000 -4.739 -1.643   
## cyl -1.508 0.415 -0.447 -3.636 0.001 -2.356 -0.660   
## ----------------------------------------------------------------------------------------  
##   
##   
##   
## No more variables to be added/removed.  
##   
##   
## Final Model Output   
## ------------------  
##   
## Model Summary   
## --------------------------------------------------------------  
## R 0.911 RMSE 2.568   
## R-Squared 0.830 Coef. Var 12.780   
## Adj. R-Squared 0.819 MSE 6.592   
## Pred R-Squared 0.790 MAE 1.921   
## --------------------------------------------------------------  
## RMSE: Root Mean Square Error   
## MSE: Mean Square Error   
## MAE: Mean Absolute Error   
##   
## ANOVA   
## --------------------------------------------------------------------  
## Sum of   
## Squares DF Mean Square F Sig.   
## --------------------------------------------------------------------  
## Regression 934.875 2 467.438 70.908 0.0000   
## Residual 191.172 29 6.592   
## Total 1126.047 31   
## --------------------------------------------------------------------  
##   
## Parameter Estimates   
## ----------------------------------------------------------------------------------------  
## model Beta Std. Error Std. Beta t Sig lower upper   
## ----------------------------------------------------------------------------------------  
## (Intercept) 39.686 1.715 23.141 0.000 36.179 43.194   
## wt -3.191 0.757 -0.518 -4.216 0.000 -4.739 -1.643   
## cyl -1.508 0.415 -0.447 -3.636 0.001 -2.356 -0.660   
## ----------------------------------------------------------------------------------------

##   
## Stepwise Selection Summary   
## -------------------------------------------------------------------------------------  
## Added/ Adj.   
## Step Variable Removed R-Square R-Square C(p) AIC RMSE   
## -------------------------------------------------------------------------------------  
## 1 wt addition 0.753 0.745 11.6270 166.0294 3.0459   
## 2 cyl addition 0.830 0.819 1.2190 156.0101 2.5675   
## -------------------------------------------------------------------------------------

### Results

The forward selection created a model consisting of three variabes; wt, cyl, and hp. The model had a RsquaredAdj of 0.826, a RMSE of 2.512, and a MAE of 1.845. The backward selection process created a model consisting of five variables; disp, hp, wt, qsec, and am. The model had a RsquaredAdj of 0.838, a RMSE of 2.429, and a MAE of 1.815. The stepwise selection process created a model consisting of two variables; wt and cyl. The model had a RsquaredAdj of 0.819, a RMSE of 2.568, and a MAE of 1.921.

### Discussion

The results of the model selection processes to determine what variables contribute the most to mpg are not entirely suprising. All three processes included wt. One would expect the weight of a vehicle to contribute to the fuel efficiency of it. The forward and stepwise selection processes both included cyl as an importand factor. This is not suprising since typically the more cylinders a vehicle engine has, the more power it creates which requires more fuel, thus lowering it’s efficiency. The forward and backward selection processes both included hp, which is not suprising since the more power a vehicle creates, the more energy it will require, likely lowering it’s efficeiency.

Model selection is not a science per se, it is an art. Depending on the experiment and the experimenter, different criteria will determine what model is best, and this may change between individuals. In this project, I will select the two variable model from the stepwise selection process that included wt and cyl. When comparing this model to the three variable model created from the forward selection process and the five variable model created from the backward selection process, they all had very similar RsquaredAdj, RMSE, and MAE. The two variable model had the fewest variables though, and when most of the model summary results are similar, I will follow the prinicple of parsimony and select the model with the fewest variables.

### Literature Cited

I used the mtcars dataset that comes with R.