HW16

109006206

Set Working Directories & Reading Files

```r
library(rpart)
setwd("/Users/olivia/Documents/Documents/Study/Semester 6/BACS/HW16")

# Load the data and remove missing values
cars <- read.table("auto-data.txt", header=FALSE, na.strings = "?")
names(cars) <- c("mpg", "cylinders", "displacement", "horsepower", "weight", "acceleration",
                 "model_year", "origin", "car_name")
cars$car_name <- NULL
cars <- na.omit(cars)
# IMPORTANT: Shuffle the rows of data in advance for this project!
set.seed(27935752)
cars <- cars[sample(1:nrow(cars)),]
# DV and IV of formulas we are interested in
set.seed(27935752)
cars_full <- mpg ~ cylinders + displacement + horsepower + weight + acceleration +
  model_year + factor(origin)
cars_reduced <- mpg ~ weight + acceleration + model_year + factor(origin)
cars_full_poly2 <- mpg ~ poly(cylinders, 2) + poly(displacement, 2) + poly(horsepower, 2) +
  poly(weight, 2) + poly(acceleration, 2) + model_year +
  factor(origin)
cars_reduced_poly2 <- mpg ~ poly(weight, 2) + poly(acceleration,2) + model_year +
  factor(origin)
cars_reduced_poly6 <- mpg ~ poly(weight, 6) + poly(acceleration,6) + model_year +
  factor(origin)
```

QUESTION 1

```r
lm_full <- lm(cars_full, data = cars)


lm_reduced <- lm(cars_reduced, data = cars)


lm_poly2_full <- lm(cars_full_poly2, data = cars)
```

```r
lm_poly2_reduced <- lm(cars_reduced_poly2, data = cars)


lm_poly6_reduced <- lm(cars_reduced_poly6, data = cars)


library(rpart)
rt_full <- rpart(formula = cars_full, data = cars)
rt_reduced <- rpart(formula = cars_reduced, data = cars)


mse_in <- function(model, data) {
  predicted <- predict(model, data)
  actual <- data$mpg
  mse <- mean((actual - predicted)^2)
  return(mse)
}


mse_lm_full <- mse_in(lm_full, cars)
# Compute MSEin for lm_reduced
mse_lm_reduced <- mse_in(lm_reduced, cars)


# Compute MSEin for lm_poly2_full
mse_lm_poly2_full <- mse_in(lm_poly2_full, cars)


# Compute MSEin for lm_poly2_reduced
mse_lm_poly2_reduced <- mse_in(lm_poly2_reduced, cars)


# Compute MSEin for lm_poly6_reduced
mse_lm_poly6_reduced <- mse_in(lm_poly6_reduced, cars)


# Compute MSEin for rt_full
mse_rt_full <- mse_in(rt_full, cars)


# Compute MSEin for rt_reduced
mse_rt_reduced <- mse_in(rt_reduced, cars)


# Create a data frame to report the MSEin values
mse_report <- data.frame(
  Model = c("lm_full", "lm_reduced", "lm_poly2_full", "lm_poly2_reduced", "lm_poly6_reduced", "rt_full"
  MSEin = c(mse_lm_full, mse_lm_reduced, mse_lm_poly2_full, mse_lm_poly2_reduced, mse_lm_poly6_reduced,
)


# Print the MSEin report
print(mse_report)
```

```
##                Model       MSEin
## 1            lm_full 10.682122
## 2         lm_reduced 10.971643
## 3      lm_poly2_full  7.919030
## 4   lm_poly2_reduced  8.364546
## 5   lm_poly6_reduced  8.254377
## 6            rt_full  9.155146
## 7         rt_reduced  9.501344
```

QUESTION 2

Part A

```
train_indices <- sample(1:nrow(cars), size=0.70*nrow(cars))
train_indices
```

```
##   [1] 133 191 240   8 286 119 347 174 165  72 254 199 357  60 355  95 384 297
##  [19]  82 263 330 337 190 118 320  11 318  81 292  43 277  93  59 160 253 210
##  [37] 198  23 260 241 178  55 264 387 214 223 334 173 383 202 303  21 349 106
##  [55]   5 271 251 283 299 157 302 193 252 149   3 339 176 259 134 182 295 209
##  [73] 212 226 228  37 184 249 389 331 103  74 104  30 380 245 227 238  46 168
##  [91] 372 112 150 269 114 388 310  84 129 311  52  57  44 375 300 338 284 213
## [109] 352 127  20  80 335 267  15 272 231   7 218  47 155 128  29   9  85 170
## [127] 151 373 342 159 344 144 122  31 294  42 385 164 386 197 340 179 235  14
## [145] 360  41  12 185 346  38 224 137 180 358  18 205 257 365 329 208  62  58
## [163]  16 138 391 367 322 186 315 194 288 196 313 316 348 336 261 307 324 270
## [181]  75 250 306 374 356 171  64 377 314 131 308  50  51 156 325 296 341  76
## [199] 370 216 363  22 220 130  97 268  56 221 247 163  39   4 289  19 248 167
## [217] 351 293  53 354 121  26 321 229 161 187 328 298 273 280  61 115 233 378
## [235] 234  67  27 332  13 147 139 169  24 188 323  77 225  96 319 353  45 236
## [253] 120 266  71   6 125  32 243 117 381 361 109 200 345 239  79 110 201 154
## [271] 145 207 305 132
```

Part B

```
train_set <- cars[train_indices,]
trained_model <- lm(lm_reduced, data = train_set)
coefficients(trained_model)
```

```
##     (Intercept)          weight     acceleration      model_year factor(origin)2
##   -15.607401085    -0.005819158    -0.036373612     0.736417738     2.391556979
## factor(origin)3
##     2.899296257
```

Part C

```
test_set <- cars[-train_indices,]
mpg_predicted <- predict(trained_model, test_set)

mse_in2 <- mse_in(trained_model,train_set)
mse_out2 <- mean((test_set$mpg - mpg_predicted)^2)

mse_in2  # In-sample mean squared error
```

```
## [1] 10.60174
```

```r
mse_out2  # Out-of-sample mean squared error
```

```
## [1] 12.27128
```

Part D

```r
mpg_actual <- test_set$mpg
pred_err <- mpg_actual - mpg_predicted
results <- data.frame(Actual = mpg_actual, Predicted = mpg_predicted, out = pred_err)

# Show the first several rows of the results data frame
head(results,5)
```

```
##     Actual Predicted        out
## 67      17  15.62843  1.3715672
## 372     29  29.50350 -0.5035027
## 158     15  13.27764  1.7223610
## 82      28  26.38139  1.6186117
## 182     33  31.44130  1.5587006
```

Question 3

Part A

Part I

```r
k_fold_mse1 <- function(dataset, k=10,model) {
  fold_pred_errors <- sapply(1:k, \(i) {
    fold_i_pe1(i, k, dataset,model)
  })
  pred_errors <- unlist(fold_pred_errors)
  mean(pred_errors^2)
}


fold_i_pe1 <- function(i, k, dataset,model) {
  folds <- cut(1:nrow(dataset), k, labels = FALSE)
  test_indices <- which(folds == i)
  test_set <- dataset[test_indices,]
  train_set = dataset[-test_indices, ]
  trained_model = lm(model,data=train_set)
  predictions = predict(trained_model, test_set)
  actuals=test_set[,1]
  pred_errors = actuals - predictions
  return(pred_errors)
}


models <- list(
  lm_full = list(formula = cars_full),
  lm_reduced = list(formula = cars_reduced),
  lm_poly2_full = list(formula = cars_full_poly2),
  lm_poly2_reduced = list(formula = cars_reduced_poly2),
  lm_poly6_reduced = list(formula = cars_reduced_poly6)
)
# Perform k-fold cross-validation for each model and report MSEout
results <- lapply(models, function(model) {
  mse_out <- k_fold_mse1(cars, k = 10,model)
  return(mse_out)
})


results_df <- data.frame(Model = names(results), MSEout = unlist(results))

k_fold_mse2 <- function(dataset, k = 10, model) {
  fold_pred_errors <- sapply(1:k, \(i) {
    fold_i_pe2(i, k, dataset, model)
```

```r
  })
  pred_errors <- unlist(fold_pred_errors)
  mean(pred_errors^2)
}


fold_i_pe2 <- function(i, k, dataset, model) {
  folds <- cut(1:nrow(dataset), k, labels = FALSE)
  test_indices <- which(folds == i)
  test_set <- dataset[test_indices,]
  train_set <- dataset[-test_indices, ]
  trained_model <- rpart(model, data = train_set)
  predictions <- predict(trained_model, test_set)
  actuals <- test_set[, 1]
  pred_errors <- actuals - predictions
  return(pred_errors)
}


models2 <- list(
  rt_full = list(formula = rt_full),
  rt_reduced = list(formula = rt_reduced)
)


results2 <- lapply(models2, function(model) {
  mse_out2 <- k_fold_mse2(cars, k = 10, model)
  return(mse_out2)
})



# Create a data frame to store the results
results_df2 <- data.frame(Model = names(results2), MSEout = unlist(results2))

MSE_Out<- rbind(results_df,results_df2)
MSE_Out
```

```
##                          Model    MSEout
## lm_full                lm_full 11.262460
## lm_reduced          lm_reduced 11.415855
## lm_poly2_full      lm_poly2_full  8.599373
## lm_poly2_reduced lm_poly2_reduced  8.818607
## lm_poly6_reduced lm_poly6_reduced  9.267369
## rt_full                rt_full 13.342221
## rt_reduced          rt_reduced 13.476272
```

Part II

```
cbind(MSE_Out[2],mse_report[2])
```

```
##                    MSEout     MSEin
## lm_full          11.262460 10.682122
## lm_reduced       11.415855 10.971643
## lm_poly2_full     8.599373  7.919030
## lm_poly2_reduced  8.818607  8.364546
## lm_poly6_reduced  9.267369  8.254377
## rt_full          13.342221  9.155146
## rt_reduced       13.476272  9.501344
```

Answer : MSE_Out is Bigger.

Part III

```
set.seed(NULL)
repetitions <- 5

mse_out_repetitions <- replicate(repetitions, {
  mse_out <- k_fold_mse1(cars, k = 10, model = lm_full)
  return(mse_out)
})

mse_out_repetitions
```

```
## [1] 11.26246 11.26246 11.26246 11.26246 11.26246
```

Part B

Part I

```
k <- 392
n <- nrow(cars)  # Assuming 'cars' is the dataset

# Initialize empty vectors to store the number of rows
train_rows <- numeric(k)
test_rows <- numeric(k)

# Calculate the number of rows in each dataset for each iteration
for (i in 1:k) {
  test_rows[i] <- ceiling(n / k)
  train_rows[i] <- n - test_rows[i]
```

```
  n <- n - test_rows[i]
}


# Print the number of rows in the training dataset and test dataset for each iteration
for (i in 1:k) {
  cat("Iteration", i, ": Train Rows =", train_rows[i], ", Test Rows =", test_rows[i], "\n")
}
```

```
## Iteration 1 : Train Rows = 391 , Test Rows = 1
## Iteration 2 : Train Rows = 390 , Test Rows = 1
## Iteration 3 : Train Rows = 389 , Test Rows = 1
## Iteration 4 : Train Rows = 388 , Test Rows = 1
## Iteration 5 : Train Rows = 387 , Test Rows = 1
## Iteration 6 : Train Rows = 386 , Test Rows = 1
## Iteration 7 : Train Rows = 385 , Test Rows = 1
## Iteration 8 : Train Rows = 384 , Test Rows = 1
## Iteration 9 : Train Rows = 383 , Test Rows = 1
## Iteration 10 : Train Rows = 382 , Test Rows = 1
## Iteration 11 : Train Rows = 381 , Test Rows = 1
## Iteration 12 : Train Rows = 380 , Test Rows = 1
## Iteration 13 : Train Rows = 379 , Test Rows = 1
## Iteration 14 : Train Rows = 378 , Test Rows = 1
## Iteration 15 : Train Rows = 377 , Test Rows = 1
## Iteration 16 : Train Rows = 376 , Test Rows = 1
## Iteration 17 : Train Rows = 375 , Test Rows = 1
## Iteration 18 : Train Rows = 374 , Test Rows = 1
## Iteration 19 : Train Rows = 373 , Test Rows = 1
## Iteration 20 : Train Rows = 372 , Test Rows = 1
## Iteration 21 : Train Rows = 371 , Test Rows = 1
## Iteration 22 : Train Rows = 370 , Test Rows = 1
## Iteration 23 : Train Rows = 369 , Test Rows = 1
## Iteration 24 : Train Rows = 368 , Test Rows = 1
## Iteration 25 : Train Rows = 367 , Test Rows = 1
## Iteration 26 : Train Rows = 366 , Test Rows = 1
## Iteration 27 : Train Rows = 365 , Test Rows = 1
## Iteration 28 : Train Rows = 364 , Test Rows = 1
## Iteration 29 : Train Rows = 363 , Test Rows = 1
## Iteration 30 : Train Rows = 362 , Test Rows = 1
## Iteration 31 : Train Rows = 361 , Test Rows = 1
## Iteration 32 : Train Rows = 360 , Test Rows = 1
## Iteration 33 : Train Rows = 359 , Test Rows = 1
## Iteration 34 : Train Rows = 358 , Test Rows = 1
## Iteration 35 : Train Rows = 357 , Test Rows = 1
```

```
## Iteration 36 : Train Rows = 356 , Test Rows = 1
## Iteration 37 : Train Rows = 355 , Test Rows = 1
## Iteration 38 : Train Rows = 354 , Test Rows = 1
## Iteration 39 : Train Rows = 353 , Test Rows = 1
## Iteration 40 : Train Rows = 352 , Test Rows = 1
## Iteration 41 : Train Rows = 351 , Test Rows = 1
## Iteration 42 : Train Rows = 350 , Test Rows = 1
## Iteration 43 : Train Rows = 349 , Test Rows = 1
## Iteration 44 : Train Rows = 348 , Test Rows = 1
## Iteration 45 : Train Rows = 347 , Test Rows = 1
## Iteration 46 : Train Rows = 346 , Test Rows = 1
## Iteration 47 : Train Rows = 345 , Test Rows = 1
## Iteration 48 : Train Rows = 344 , Test Rows = 1
## Iteration 49 : Train Rows = 343 , Test Rows = 1
## Iteration 50 : Train Rows = 342 , Test Rows = 1
## Iteration 51 : Train Rows = 341 , Test Rows = 1
## Iteration 52 : Train Rows = 340 , Test Rows = 1
## Iteration 53 : Train Rows = 339 , Test Rows = 1
## Iteration 54 : Train Rows = 338 , Test Rows = 1
## Iteration 55 : Train Rows = 337 , Test Rows = 1
## Iteration 56 : Train Rows = 336 , Test Rows = 1
## Iteration 57 : Train Rows = 335 , Test Rows = 1
## Iteration 58 : Train Rows = 334 , Test Rows = 1
## Iteration 59 : Train Rows = 333 , Test Rows = 1
## Iteration 60 : Train Rows = 332 , Test Rows = 1
## Iteration 61 : Train Rows = 331 , Test Rows = 1
## Iteration 62 : Train Rows = 330 , Test Rows = 1
## Iteration 63 : Train Rows = 329 , Test Rows = 1
## Iteration 64 : Train Rows = 328 , Test Rows = 1
## Iteration 65 : Train Rows = 327 , Test Rows = 1
## Iteration 66 : Train Rows = 326 , Test Rows = 1
## Iteration 67 : Train Rows = 325 , Test Rows = 1
## Iteration 68 : Train Rows = 324 , Test Rows = 1
## Iteration 69 : Train Rows = 323 , Test Rows = 1
## Iteration 70 : Train Rows = 322 , Test Rows = 1
## Iteration 71 : Train Rows = 321 , Test Rows = 1
## Iteration 72 : Train Rows = 320 , Test Rows = 1
## Iteration 73 : Train Rows = 319 , Test Rows = 1
## Iteration 74 : Train Rows = 318 , Test Rows = 1
## Iteration 75 : Train Rows = 317 , Test Rows = 1
## Iteration 76 : Train Rows = 316 , Test Rows = 1
## Iteration 77 : Train Rows = 315 , Test Rows = 1
## Iteration 78 : Train Rows = 314 , Test Rows = 1
```

```
## Iteration 79 : Train Rows = 313 , Test Rows = 1
## Iteration 80 : Train Rows = 312 , Test Rows = 1
## Iteration 81 : Train Rows = 311 , Test Rows = 1
## Iteration 82 : Train Rows = 310 , Test Rows = 1
## Iteration 83 : Train Rows = 309 , Test Rows = 1
## Iteration 84 : Train Rows = 308 , Test Rows = 1
## Iteration 85 : Train Rows = 307 , Test Rows = 1
## Iteration 86 : Train Rows = 306 , Test Rows = 1
## Iteration 87 : Train Rows = 305 , Test Rows = 1
## Iteration 88 : Train Rows = 304 , Test Rows = 1
## Iteration 89 : Train Rows = 303 , Test Rows = 1
## Iteration 90 : Train Rows = 302 , Test Rows = 1
## Iteration 91 : Train Rows = 301 , Test Rows = 1
## Iteration 92 : Train Rows = 300 , Test Rows = 1
## Iteration 93 : Train Rows = 299 , Test Rows = 1
## Iteration 94 : Train Rows = 298 , Test Rows = 1
## Iteration 95 : Train Rows = 297 , Test Rows = 1
## Iteration 96 : Train Rows = 296 , Test Rows = 1
## Iteration 97 : Train Rows = 295 , Test Rows = 1
## Iteration 98 : Train Rows = 294 , Test Rows = 1
## Iteration 99 : Train Rows = 293 , Test Rows = 1
## Iteration 100 : Train Rows = 292 , Test Rows = 1
## Iteration 101 : Train Rows = 291 , Test Rows = 1
## Iteration 102 : Train Rows = 290 , Test Rows = 1
## Iteration 103 : Train Rows = 289 , Test Rows = 1
## Iteration 104 : Train Rows = 288 , Test Rows = 1
## Iteration 105 : Train Rows = 287 , Test Rows = 1
## Iteration 106 : Train Rows = 286 , Test Rows = 1
## Iteration 107 : Train Rows = 285 , Test Rows = 1
## Iteration 108 : Train Rows = 284 , Test Rows = 1
## Iteration 109 : Train Rows = 283 , Test Rows = 1
## Iteration 110 : Train Rows = 282 , Test Rows = 1
## Iteration 111 : Train Rows = 281 , Test Rows = 1
## Iteration 112 : Train Rows = 280 , Test Rows = 1
## Iteration 113 : Train Rows = 279 , Test Rows = 1
## Iteration 114 : Train Rows = 278 , Test Rows = 1
## Iteration 115 : Train Rows = 277 , Test Rows = 1
## Iteration 116 : Train Rows = 276 , Test Rows = 1
## Iteration 117 : Train Rows = 275 , Test Rows = 1
## Iteration 118 : Train Rows = 274 , Test Rows = 1
## Iteration 119 : Train Rows = 273 , Test Rows = 1
## Iteration 120 : Train Rows = 272 , Test Rows = 1
## Iteration 121 : Train Rows = 271 , Test Rows = 1
```

```
## Iteration 122 : Train Rows = 270 , Test Rows = 1
## Iteration 123 : Train Rows = 269 , Test Rows = 1
## Iteration 124 : Train Rows = 268 , Test Rows = 1
## Iteration 125 : Train Rows = 267 , Test Rows = 1
## Iteration 126 : Train Rows = 266 , Test Rows = 1
## Iteration 127 : Train Rows = 265 , Test Rows = 1
## Iteration 128 : Train Rows = 264 , Test Rows = 1
## Iteration 129 : Train Rows = 263 , Test Rows = 1
## Iteration 130 : Train Rows = 262 , Test Rows = 1
## Iteration 131 : Train Rows = 261 , Test Rows = 1
## Iteration 132 : Train Rows = 260 , Test Rows = 1
## Iteration 133 : Train Rows = 259 , Test Rows = 1
## Iteration 134 : Train Rows = 258 , Test Rows = 1
## Iteration 135 : Train Rows = 257 , Test Rows = 1
## Iteration 136 : Train Rows = 256 , Test Rows = 1
## Iteration 137 : Train Rows = 255 , Test Rows = 1
## Iteration 138 : Train Rows = 254 , Test Rows = 1
## Iteration 139 : Train Rows = 253 , Test Rows = 1
## Iteration 140 : Train Rows = 252 , Test Rows = 1
## Iteration 141 : Train Rows = 251 , Test Rows = 1
## Iteration 142 : Train Rows = 250 , Test Rows = 1
## Iteration 143 : Train Rows = 249 , Test Rows = 1
## Iteration 144 : Train Rows = 248 , Test Rows = 1
## Iteration 145 : Train Rows = 247 , Test Rows = 1
## Iteration 146 : Train Rows = 246 , Test Rows = 1
## Iteration 147 : Train Rows = 245 , Test Rows = 1
## Iteration 148 : Train Rows = 244 , Test Rows = 1
## Iteration 149 : Train Rows = 243 , Test Rows = 1
## Iteration 150 : Train Rows = 242 , Test Rows = 1
## Iteration 151 : Train Rows = 241 , Test Rows = 1
## Iteration 152 : Train Rows = 240 , Test Rows = 1
## Iteration 153 : Train Rows = 239 , Test Rows = 1
## Iteration 154 : Train Rows = 238 , Test Rows = 1
## Iteration 155 : Train Rows = 237 , Test Rows = 1
## Iteration 156 : Train Rows = 236 , Test Rows = 1
## Iteration 157 : Train Rows = 235 , Test Rows = 1
## Iteration 158 : Train Rows = 234 , Test Rows = 1
## Iteration 159 : Train Rows = 233 , Test Rows = 1
## Iteration 160 : Train Rows = 232 , Test Rows = 1
## Iteration 161 : Train Rows = 231 , Test Rows = 1
## Iteration 162 : Train Rows = 230 , Test Rows = 1
## Iteration 163 : Train Rows = 229 , Test Rows = 1
## Iteration 164 : Train Rows = 228 , Test Rows = 1
```

```
## Iteration 165 : Train Rows = 227 , Test Rows = 1
## Iteration 166 : Train Rows = 226 , Test Rows = 1
## Iteration 167 : Train Rows = 225 , Test Rows = 1
## Iteration 168 : Train Rows = 224 , Test Rows = 1
## Iteration 169 : Train Rows = 223 , Test Rows = 1
## Iteration 170 : Train Rows = 222 , Test Rows = 1
## Iteration 171 : Train Rows = 221 , Test Rows = 1
## Iteration 172 : Train Rows = 220 , Test Rows = 1
## Iteration 173 : Train Rows = 219 , Test Rows = 1
## Iteration 174 : Train Rows = 218 , Test Rows = 1
## Iteration 175 : Train Rows = 217 , Test Rows = 1
## Iteration 176 : Train Rows = 216 , Test Rows = 1
## Iteration 177 : Train Rows = 215 , Test Rows = 1
## Iteration 178 : Train Rows = 214 , Test Rows = 1
## Iteration 179 : Train Rows = 213 , Test Rows = 1
## Iteration 180 : Train Rows = 212 , Test Rows = 1
## Iteration 181 : Train Rows = 211 , Test Rows = 1
## Iteration 182 : Train Rows = 210 , Test Rows = 1
## Iteration 183 : Train Rows = 209 , Test Rows = 1
## Iteration 184 : Train Rows = 208 , Test Rows = 1
## Iteration 185 : Train Rows = 207 , Test Rows = 1
## Iteration 186 : Train Rows = 206 , Test Rows = 1
## Iteration 187 : Train Rows = 205 , Test Rows = 1
## Iteration 188 : Train Rows = 204 , Test Rows = 1
## Iteration 189 : Train Rows = 203 , Test Rows = 1
## Iteration 190 : Train Rows = 202 , Test Rows = 1
## Iteration 191 : Train Rows = 201 , Test Rows = 1
## Iteration 192 : Train Rows = 200 , Test Rows = 1
## Iteration 193 : Train Rows = 199 , Test Rows = 1
## Iteration 194 : Train Rows = 198 , Test Rows = 1
## Iteration 195 : Train Rows = 197 , Test Rows = 1
## Iteration 196 : Train Rows = 196 , Test Rows = 1
## Iteration 197 : Train Rows = 195 , Test Rows = 1
## Iteration 198 : Train Rows = 194 , Test Rows = 1
## Iteration 199 : Train Rows = 193 , Test Rows = 1
## Iteration 200 : Train Rows = 192 , Test Rows = 1
## Iteration 201 : Train Rows = 191 , Test Rows = 1
## Iteration 202 : Train Rows = 190 , Test Rows = 1
## Iteration 203 : Train Rows = 189 , Test Rows = 1
## Iteration 204 : Train Rows = 188 , Test Rows = 1
## Iteration 205 : Train Rows = 187 , Test Rows = 1
## Iteration 206 : Train Rows = 186 , Test Rows = 1
## Iteration 207 : Train Rows = 185 , Test Rows = 1
```

```
## Iteration 208 : Train Rows = 184 , Test Rows = 1
## Iteration 209 : Train Rows = 183 , Test Rows = 1
## Iteration 210 : Train Rows = 182 , Test Rows = 1
## Iteration 211 : Train Rows = 181 , Test Rows = 1
## Iteration 212 : Train Rows = 180 , Test Rows = 1
## Iteration 213 : Train Rows = 179 , Test Rows = 1
## Iteration 214 : Train Rows = 178 , Test Rows = 1
## Iteration 215 : Train Rows = 177 , Test Rows = 1
## Iteration 216 : Train Rows = 176 , Test Rows = 1
## Iteration 217 : Train Rows = 175 , Test Rows = 1
## Iteration 218 : Train Rows = 174 , Test Rows = 1
## Iteration 219 : Train Rows = 173 , Test Rows = 1
## Iteration 220 : Train Rows = 172 , Test Rows = 1
## Iteration 221 : Train Rows = 171 , Test Rows = 1
## Iteration 222 : Train Rows = 170 , Test Rows = 1
## Iteration 223 : Train Rows = 169 , Test Rows = 1
## Iteration 224 : Train Rows = 168 , Test Rows = 1
## Iteration 225 : Train Rows = 167 , Test Rows = 1
## Iteration 226 : Train Rows = 166 , Test Rows = 1
## Iteration 227 : Train Rows = 165 , Test Rows = 1
## Iteration 228 : Train Rows = 164 , Test Rows = 1
## Iteration 229 : Train Rows = 163 , Test Rows = 1
## Iteration 230 : Train Rows = 162 , Test Rows = 1
## Iteration 231 : Train Rows = 161 , Test Rows = 1
## Iteration 232 : Train Rows = 160 , Test Rows = 1
## Iteration 233 : Train Rows = 159 , Test Rows = 1
## Iteration 234 : Train Rows = 158 , Test Rows = 1
## Iteration 235 : Train Rows = 157 , Test Rows = 1
## Iteration 236 : Train Rows = 156 , Test Rows = 1
## Iteration 237 : Train Rows = 155 , Test Rows = 1
## Iteration 238 : Train Rows = 154 , Test Rows = 1
## Iteration 239 : Train Rows = 153 , Test Rows = 1
## Iteration 240 : Train Rows = 152 , Test Rows = 1
## Iteration 241 : Train Rows = 151 , Test Rows = 1
## Iteration 242 : Train Rows = 150 , Test Rows = 1
## Iteration 243 : Train Rows = 149 , Test Rows = 1
## Iteration 244 : Train Rows = 148 , Test Rows = 1
## Iteration 245 : Train Rows = 147 , Test Rows = 1
## Iteration 246 : Train Rows = 146 , Test Rows = 1
## Iteration 247 : Train Rows = 145 , Test Rows = 1
## Iteration 248 : Train Rows = 144 , Test Rows = 1
## Iteration 249 : Train Rows = 143 , Test Rows = 1
## Iteration 250 : Train Rows = 142 , Test Rows = 1
```

```
## Iteration 251 : Train Rows = 141 , Test Rows = 1
## Iteration 252 : Train Rows = 140 , Test Rows = 1
## Iteration 253 : Train Rows = 139 , Test Rows = 1
## Iteration 254 : Train Rows = 138 , Test Rows = 1
## Iteration 255 : Train Rows = 137 , Test Rows = 1
## Iteration 256 : Train Rows = 136 , Test Rows = 1
## Iteration 257 : Train Rows = 135 , Test Rows = 1
## Iteration 258 : Train Rows = 134 , Test Rows = 1
## Iteration 259 : Train Rows = 133 , Test Rows = 1
## Iteration 260 : Train Rows = 132 , Test Rows = 1
## Iteration 261 : Train Rows = 131 , Test Rows = 1
## Iteration 262 : Train Rows = 130 , Test Rows = 1
## Iteration 263 : Train Rows = 129 , Test Rows = 1
## Iteration 264 : Train Rows = 128 , Test Rows = 1
## Iteration 265 : Train Rows = 127 , Test Rows = 1
## Iteration 266 : Train Rows = 126 , Test Rows = 1
## Iteration 267 : Train Rows = 125 , Test Rows = 1
## Iteration 268 : Train Rows = 124 , Test Rows = 1
## Iteration 269 : Train Rows = 123 , Test Rows = 1
## Iteration 270 : Train Rows = 122 , Test Rows = 1
## Iteration 271 : Train Rows = 121 , Test Rows = 1
## Iteration 272 : Train Rows = 120 , Test Rows = 1
## Iteration 273 : Train Rows = 119 , Test Rows = 1
## Iteration 274 : Train Rows = 118 , Test Rows = 1
## Iteration 275 : Train Rows = 117 , Test Rows = 1
## Iteration 276 : Train Rows = 116 , Test Rows = 1
## Iteration 277 : Train Rows = 115 , Test Rows = 1
## Iteration 278 : Train Rows = 114 , Test Rows = 1
## Iteration 279 : Train Rows = 113 , Test Rows = 1
## Iteration 280 : Train Rows = 112 , Test Rows = 1
## Iteration 281 : Train Rows = 111 , Test Rows = 1
## Iteration 282 : Train Rows = 110 , Test Rows = 1
## Iteration 283 : Train Rows = 109 , Test Rows = 1
## Iteration 284 : Train Rows = 108 , Test Rows = 1
## Iteration 285 : Train Rows = 107 , Test Rows = 1
## Iteration 286 : Train Rows = 106 , Test Rows = 1
## Iteration 287 : Train Rows = 105 , Test Rows = 1
## Iteration 288 : Train Rows = 104 , Test Rows = 1
## Iteration 289 : Train Rows = 103 , Test Rows = 1
## Iteration 290 : Train Rows = 102 , Test Rows = 1
## Iteration 291 : Train Rows = 101 , Test Rows = 1
## Iteration 292 : Train Rows = 100 , Test Rows = 1
## Iteration 293 : Train Rows = 99  , Test Rows = 1
```

```
## Iteration 294 : Train Rows = 98 , Test Rows = 1
## Iteration 295 : Train Rows = 97 , Test Rows = 1
## Iteration 296 : Train Rows = 96 , Test Rows = 1
## Iteration 297 : Train Rows = 95 , Test Rows = 1
## Iteration 298 : Train Rows = 94 , Test Rows = 1
## Iteration 299 : Train Rows = 93 , Test Rows = 1
## Iteration 300 : Train Rows = 92 , Test Rows = 1
## Iteration 301 : Train Rows = 91 , Test Rows = 1
## Iteration 302 : Train Rows = 90 , Test Rows = 1
## Iteration 303 : Train Rows = 89 , Test Rows = 1
## Iteration 304 : Train Rows = 88 , Test Rows = 1
## Iteration 305 : Train Rows = 87 , Test Rows = 1
## Iteration 306 : Train Rows = 86 , Test Rows = 1
## Iteration 307 : Train Rows = 85 , Test Rows = 1
## Iteration 308 : Train Rows = 84 , Test Rows = 1
## Iteration 309 : Train Rows = 83 , Test Rows = 1
## Iteration 310 : Train Rows = 82 , Test Rows = 1
## Iteration 311 : Train Rows = 81 , Test Rows = 1
## Iteration 312 : Train Rows = 80 , Test Rows = 1
## Iteration 313 : Train Rows = 79 , Test Rows = 1
## Iteration 314 : Train Rows = 78 , Test Rows = 1
## Iteration 315 : Train Rows = 77 , Test Rows = 1
## Iteration 316 : Train Rows = 76 , Test Rows = 1
## Iteration 317 : Train Rows = 75 , Test Rows = 1
## Iteration 318 : Train Rows = 74 , Test Rows = 1
## Iteration 319 : Train Rows = 73 , Test Rows = 1
## Iteration 320 : Train Rows = 72 , Test Rows = 1
## Iteration 321 : Train Rows = 71 , Test Rows = 1
## Iteration 322 : Train Rows = 70 , Test Rows = 1
## Iteration 323 : Train Rows = 69 , Test Rows = 1
## Iteration 324 : Train Rows = 68 , Test Rows = 1
## Iteration 325 : Train Rows = 67 , Test Rows = 1
## Iteration 326 : Train Rows = 66 , Test Rows = 1
## Iteration 327 : Train Rows = 65 , Test Rows = 1
## Iteration 328 : Train Rows = 64 , Test Rows = 1
## Iteration 329 : Train Rows = 63 , Test Rows = 1
## Iteration 330 : Train Rows = 62 , Test Rows = 1
## Iteration 331 : Train Rows = 61 , Test Rows = 1
## Iteration 332 : Train Rows = 60 , Test Rows = 1
## Iteration 333 : Train Rows = 59 , Test Rows = 1
## Iteration 334 : Train Rows = 58 , Test Rows = 1
## Iteration 335 : Train Rows = 57 , Test Rows = 1
## Iteration 336 : Train Rows = 56 , Test Rows = 1
```

```
## Iteration 337 : Train Rows = 55 , Test Rows = 1
## Iteration 338 : Train Rows = 54 , Test Rows = 1
## Iteration 339 : Train Rows = 53 , Test Rows = 1
## Iteration 340 : Train Rows = 52 , Test Rows = 1
## Iteration 341 : Train Rows = 51 , Test Rows = 1
## Iteration 342 : Train Rows = 50 , Test Rows = 1
## Iteration 343 : Train Rows = 49 , Test Rows = 1
## Iteration 344 : Train Rows = 48 , Test Rows = 1
## Iteration 345 : Train Rows = 47 , Test Rows = 1
## Iteration 346 : Train Rows = 46 , Test Rows = 1
## Iteration 347 : Train Rows = 45 , Test Rows = 1
## Iteration 348 : Train Rows = 44 , Test Rows = 1
## Iteration 349 : Train Rows = 43 , Test Rows = 1
## Iteration 350 : Train Rows = 42 , Test Rows = 1
## Iteration 351 : Train Rows = 41 , Test Rows = 1
## Iteration 352 : Train Rows = 40 , Test Rows = 1
## Iteration 353 : Train Rows = 39 , Test Rows = 1
## Iteration 354 : Train Rows = 38 , Test Rows = 1
## Iteration 355 : Train Rows = 37 , Test Rows = 1
## Iteration 356 : Train Rows = 36 , Test Rows = 1
## Iteration 357 : Train Rows = 35 , Test Rows = 1
## Iteration 358 : Train Rows = 34 , Test Rows = 1
## Iteration 359 : Train Rows = 33 , Test Rows = 1
## Iteration 360 : Train Rows = 32 , Test Rows = 1
## Iteration 361 : Train Rows = 31 , Test Rows = 1
## Iteration 362 : Train Rows = 30 , Test Rows = 1
## Iteration 363 : Train Rows = 29 , Test Rows = 1
## Iteration 364 : Train Rows = 28 , Test Rows = 1
## Iteration 365 : Train Rows = 27 , Test Rows = 1
## Iteration 366 : Train Rows = 26 , Test Rows = 1
## Iteration 367 : Train Rows = 25 , Test Rows = 1
## Iteration 368 : Train Rows = 24 , Test Rows = 1
## Iteration 369 : Train Rows = 23 , Test Rows = 1
## Iteration 370 : Train Rows = 22 , Test Rows = 1
## Iteration 371 : Train Rows = 21 , Test Rows = 1
## Iteration 372 : Train Rows = 20 , Test Rows = 1
## Iteration 373 : Train Rows = 19 , Test Rows = 1
## Iteration 374 : Train Rows = 18 , Test Rows = 1
## Iteration 375 : Train Rows = 17 , Test Rows = 1
## Iteration 376 : Train Rows = 16 , Test Rows = 1
## Iteration 377 : Train Rows = 15 , Test Rows = 1
## Iteration 378 : Train Rows = 14 , Test Rows = 1
## Iteration 379 : Train Rows = 13 , Test Rows = 1
```

```
## Iteration 380 : Train Rows = 12 , Test Rows = 1
## Iteration 381 : Train Rows = 11 , Test Rows = 1
## Iteration 382 : Train Rows = 10 , Test Rows = 1
## Iteration 383 : Train Rows = 9 , Test Rows = 1
## Iteration 384 : Train Rows = 8 , Test Rows = 1
## Iteration 385 : Train Rows = 7 , Test Rows = 1
## Iteration 386 : Train Rows = 6 , Test Rows = 1
## Iteration 387 : Train Rows = 5 , Test Rows = 1
## Iteration 388 : Train Rows = 4 , Test Rows = 1
## Iteration 389 : Train Rows = 3 , Test Rows = 1
## Iteration 390 : Train Rows = 2 , Test Rows = 1
## Iteration 391 : Train Rows = 1 , Test Rows = 1
## Iteration 392 : Train Rows = 0 , Test Rows = 1
```

Answer: There are 391 rows are in the training dataset and test dataset of each iteration of k-fold CV when k=392

Part II

```r
results3 <- lapply(models, function(model) {
  mse_out3 <- k_fold_mse2(cars, k = 392, model)
  return(mse_out3)
})


results_df3 <- data.frame(Model = names(results3), MSEout = unlist(results3))
# Create a data frame to store the results
results4 <- lapply(models2, function(model) {
  mse_out4 <- k_fold_mse1(cars, k = 392, model)
  return(mse_out4)
})


results_df4 <- data.frame(Model = names(results4), MSEout = unlist(results4))


MSE_Out2<- rbind(results_df3,results_df4)
MSE_Out2
```

```
##                            Model    MSEout
## lm_full                  lm_full 12.76979
## lm_reduced            lm_reduced 13.14515
## lm_poly2_full      lm_poly2_full 12.76849
## lm_poly2_reduced lm_poly2_reduced 13.43981
## lm_poly6_reduced lm_poly6_reduced 13.40653
```

```
## rt_full              rt_full 11.29344
## rt_reduced       rt_reduced 11.38004
```

Part III

```
set.seed(27935752)
repetitions <- 5

mse_out_repetitions <- replicate(repetitions, {
  mse_out <- k_fold_mse1(cars, k = 392, model = lm_full)
  return(mse_out)
})

mse_out_repetitions
```

```
## [1] 11.29344 11.29344 11.29344 11.29344 11.29344
```