# BE7023 Homework 4

*Mike Lape*

*Octobe 3, 2018*

```r
#setwd("C:/Users/lapt3u/Box/UC/Fall_2018/BE7023_Adv_Biostats/adv_biostats/hw_4")
library(car)
library(MASS)
library(leaps)
dat <- Highway1
```

1. Determine the dimension of the data.
   Show the top ten rows of the data.
   Obtain summary statistics of the data.

```r
dim(dat)
```

```
## [1] 39 12
```

```r
# The Highway1 dataset has 39 rows/observations and 12 columns/variables

# Top 10 rows of Highway1:
head(dat, 10)
```

```
##    rate   len adt trks      sigs1 slim shld lane acpt  itg lwid htype
## 1  4.58  4.99  69    8 0.20040080   55   10    8  4.6 1.20   12   FAI
## 2  2.86 16.11  73    8 0.06207325   60   10    4  4.4 1.43   12   FAI
## 3  3.02  9.75  49   10 0.10256410   60   10    4  4.7 1.54   12   FAI
## 4  2.29 10.65  61   13 0.09389671   65   10    6  3.8 0.94   12   FAI
## 5  1.61 20.01  28   12 0.04997501   70   10    4  2.2 0.65   12   FAI
## 6  6.87  5.97  30    6 2.00750419   55   10    4 24.8 0.34   12    PA
## 7  3.85  8.57  46    8 0.81668611   55    8    4 11.0 0.47   12    PA
## 8  6.12  5.24  25    9 0.57083969   55   10    4 18.5 0.38   12    PA
## 9  3.29 15.79  43   12 1.45333122   50    4    4  7.5 0.95   12    PA
## 10 5.88  8.26  23    7 1.33106538   50    5    4  8.2 0.12   12    PA
```

```r
# Summary statistics of Highway1:
summary(dat)
```

```
##       rate            len              adt             trks
##  Min.   :1.610   Min.   : 2.960   Min.   : 1.00   Min.   : 6.000
##  1st Qu.:2.630   1st Qu.: 7.995   1st Qu.: 5.00   1st Qu.: 8.000
##  Median :3.050   Median :11.390   Median :13.00   Median : 9.000
##  Mean   :3.933   Mean   :12.884   Mean   :19.62   Mean   : 9.333
##  3rd Qu.:4.595   3rd Qu.:17.800   3rd Qu.:24.00   3rd Qu.:11.000
```

```
##  Max.   :9.230   Max.   :40.090   Max.   :73.00   Max.   :15.000
##      sigs1              slim         shld            lane
##  Min.   :0.04545   Min.   :40   Min.   : 1.000   Min.   :2.000
##  1st Qu.:0.08738   1st Qu.:50   1st Qu.: 4.000   1st Qu.:2.000
##  Median :0.17666   Median :55   Median : 8.000   Median :2.000
##  Mean   :0.51072   Mean   :55   Mean   : 6.872   Mean   :3.128
##  3rd Qu.:0.71515   3rd Qu.:60   3rd Qu.:10.000   3rd Qu.:4.000
##  Max.   :2.78933   Max.   :70   Max.   :10.000   Max.   :8.000
##       acpt            itg             lwid        htype
##  Min.   : 2.20   Min.   :0.0000   Min.   :10.00   FAI: 5
##  1st Qu.: 6.95   1st Qu.:0.0000   1st Qu.:12.00   MA :13
##  Median :10.30   Median :0.1300   Median :12.00   MC : 2
##  Mean   :12.16   Mean   :0.2964   Mean   :11.95   PA :19
##  3rd Qu.:14.60   3rd Qu.:0.3600   3rd Qu.:12.00
##  Max.   :53.00   Max.   :1.5400   Max.   :13.00
```

2. Describe the data.

```r
# This dataset has auto accident rates per million vehicle miles for 39 stetches of road
# in Minnesota in the year 1973, as well as other variables to describe that stretch of
# road.  It was put together by Carl Hoffstedt.
# Variables:
#   rate:   auto accident rate in 1973 [accidents/million vehicle miles]
#   len:    length of a highway segment [miles]
#   adt:    average daily traffic count in thousands
#   trks:   truck volume as percent of total volume
#   sigs1:  number of signals per mile of roadway
#   slim:   speed limit of stretch of road in 1973
#   shld:   width of outer shoulder on road [feet]
#   lane:   total number of traffic lanes
#   acpt:   nummber of access points per mile
#   itg:    number of freeway-type interchanges per mile
#   lwid:   lane width [feet]
#   htype:  Type of roadway or source of funding for road
sapply(dat, class)
```

```
##      rate       len       adt      trks     sigs1      slim      shld
## "numeric" "numeric" "integer" "integer" "numeric" "integer" "integer"
##      lane      acpt       itg      lwid     htype
## "integer" "numeric" "numeric" "integer"  "factor"
```

```r
levels(dat$htype)
```

```
## [1] "FAI" "MA"  "MC"  "PA"
```

```r
# All variabes are either numeric or integers (no decimals) except for htype
# which is a factor/categorical with 4 levels, "FAI", "MA", "MC", "PA".
```

3.The response variable is ‘rate.’
log2(rate) is taken to be the response variable.
The predictors are taken to be: log2(len); log2(ADT); log2(trks); log2(sigs1); slim; shld; lane; acpt; itg; lwid; hwy. (The last predictor is categorical with four levels.)
You respect the transformations recommended.
Fit a full regression model.
Comment on the output including R2.
Write the prediction model.
Identify the significant predictors.
Explain how the categorical variable ‘hwy’ is handled.
Estimate the standard deviation of the error.

```r
# Construct seperate df for holding our log transformed dataset

l_dat <-  dat[,6:12]

# Transforming variables as needed
l_dat$l_rate  <- log2(dat$rate)
l_dat$l_len   <- log2(dat$len)
l_dat$l_adt   <- log2(dat$adt)
l_dat$l_trks  <- log2(dat$trks)
l_dat$l_sigs1 <- log2(dat$sigs1)

mod_full <- lm(l_rate ~ . , l_dat)
summary(mod_full)
```

```
##
## Call:
## lm(formula = l_rate ~ ., data = l_dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.64635 -0.14705 -0.00998  0.17645  0.60761
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.047344   2.623516   2.305   0.0297 *
## slim        -0.039327   0.024236  -1.623   0.1172
## shld         0.004291   0.049281   0.087   0.9313
## lane        -0.016061   0.082264  -0.195   0.8468
## acpt         0.008727   0.011687   0.747   0.4622
## itg          0.051536   0.350312   0.147   0.8842
## lwid         0.060769   0.197391   0.308   0.7607
## htypeMA     -0.550063   0.515724  -1.067   0.2964
## htypeMC     -0.342705   0.576821  -0.594   0.5578
## htypePA     -0.755001   0.418441  -1.804   0.0832 .
## l_len       -0.214470   0.099986  -2.145   0.0419 *
## l_adt       -0.154625   0.111893  -1.382   0.1792
## l_trks      -0.197560   0.239812  -0.824   0.4178
## l_sigs1      0.192322   0.075367   2.552   0.0172 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3761 on 25 degrees of freedom
```

```
## Multiple R-squared:  0.7913, Adjusted R-squared:  0.6828
## F-statistic: 7.293 on 13 and 25 DF,  p-value: 1.247e-05
```

Our model using all predictors has an R2 of 0.683 with a p-value of 1.247e-05, indicating a decent and significant fit. We have 2 significant predictors not counting the intercept, Log2(len) and Log2(sigs1).

Prediction Equation:
Log2(rate) = 6.047 - (0.214 * Log2(len)) - (0.155 * Log2(adt)) - (0.198 * Log2(trks)) +
(0.192 * Log2(sigs1)) - (0.039 * slim) + (0.004 * shld) -
(0.016 * lane) + (0.009 * acpt)+ (0.052 * itg) + (0.061 * lwid) -
(0.550 * htype[MA]) - (0.343 * htype[MC]) - (0.755 * htype[PA])

Significant Predictors:
Log2(len) [p_val = 0.0419]: Log2 of the length of road segment in miles
Log2(sigs1) [p_val = 0.0172]: Log2 of number of signals per mile of roadway

The categorical variable htype that has 4 levels is broken down into dummy variables essentially from 1 variable creating 3 dichotomous variables one for each level except one is considered the baseline, in this case htype[FAI], so for a road with hytpe of MA it would have a htype[MA] = 1, but an htype[MC] = 0 and an htype[PA] also equal to 0, and if the htype was FAI then htype[MA,MC,PA] would all equal 0.

```r
# Standard deviation of the error, aka standard error is given as sigma
# from the lm summary, ours is: 0.376
round(summary(mod_full)$sigma,3)
```

```
## [1] 0.376
```

4. Employ the 'backward elimination procedure' on the model in Q3 to obtain a tight model according to the Akaike Information Criterion.
   Write the final prediction model.
   Compare its R2 with the R2 of the full model.
   Compare the estimates of the standard deviation of the error terms.

```r
# For backwards we will just start with our full model created previously.
tight_mod <- stepAIC(mod_full, direction = "backward")
```

```
## Start:  AIC=-65.61
## l_rate ~ slim + shld + lane + acpt + itg + lwid + htype + l_len +
##     l_adt + l_trks + l_sigs1
##
##           Df Sum of Sq    RSS     AIC
## - shld     1   0.00107 3.5380 -67.600
## - itg      1   0.00306 3.5400 -67.578
## - lane     1   0.00539 3.5424 -67.552
## - lwid     1   0.01341 3.5504 -67.464
## - acpt     1   0.07889 3.6159 -66.751
## - l_trks   1   0.09602 3.6330 -66.567
## <none>                 3.5370 -65.611
## - htype    3   0.62534 4.1623 -65.262
## - l_adt    1   0.27017 3.8071 -64.741
## - slim     1   0.37251 3.9095 -63.706
```

```
## - l_len    1   0.65095 4.1879 -61.023
## - l_sigs1  1   0.92127 4.4582 -58.584
##
## Step:  AIC=-67.6
## l_rate ~ slim + lane + acpt + itg + lwid + htype + l_len + l_adt +
##     l_trks + l_sigs1
##
##           Df Sum of Sq    RSS     AIC
## - itg      1   0.00276 3.5408 -69.569
## - lane     1   0.00571 3.5437 -69.537
## - lwid     1   0.01489 3.5529 -69.436
## - acpt     1   0.09737 3.6354 -68.541
## - l_trks   1   0.11580 3.6538 -68.344
## <none>                 3.5380 -67.600
## - htype    3   0.68361 4.2216 -66.710
## - l_adt    1   0.28417 3.8222 -66.587
## - slim     1   0.70869 4.2467 -62.479
## - l_len    1   0.72660 4.2646 -62.315
## - l_sigs1  1   1.00199 4.5400 -59.875
##
## Step:  AIC=-69.57
## l_rate ~ slim + lane + acpt + lwid + htype + l_len + l_adt +
##     l_trks + l_sigs1
##
##           Df Sum of Sq    RSS     AIC
## - lane     1   0.00516 3.5460 -71.512
## - lwid     1   0.01395 3.5547 -71.416
## - acpt     1   0.09477 3.6356 -70.539
## - l_trks   1   0.11818 3.6590 -70.289
## <none>                 3.5408 -69.569
## - l_adt    1   0.32303 3.8638 -68.164
## - htype    3   1.16490 4.7057 -64.477
## - l_len    1   0.74336 4.2842 -64.137
## - slim     1   0.78988 4.3307 -63.716
## - l_sigs1  1   1.00147 4.5423 -61.855
##
## Step:  AIC=-71.51
## l_rate ~ slim + acpt + lwid + htype + l_len + l_adt + l_trks +
##     l_sigs1
##
##           Df Sum of Sq    RSS     AIC
## - lwid     1   0.01644 3.5624 -73.332
## - acpt     1   0.09770 3.6437 -72.452
## - l_trks   1   0.11412 3.6601 -72.277
## <none>                 3.5460 -71.512
## - l_adt    1   0.38177 3.9277 -69.525
## - l_len    1   0.75404 4.3000 -65.993
## - slim     1   0.80028 4.3462 -65.576
## - htype    3   1.27572 4.8217 -65.527
## - l_sigs1  1   1.01854 4.5645 -63.665
##
## Step:  AIC=-73.33
## l_rate ~ slim + acpt + htype + l_len + l_adt + l_trks + l_sigs1
##
```

```
##           Df Sum of Sq    RSS     AIC
## - acpt     1   0.10444 3.6668 -74.205
## - l_trks   1   0.12892 3.6913 -73.946
## <none>                 3.5624 -73.332
## - l_adt    1   0.36598 3.9284 -71.518
## - slim     1   0.79675 4.3591 -67.460
## - htype    3   1.27274 4.8351 -67.419
## - l_len    1   0.86211 4.4245 -66.880
## - l_sigs1  1   1.02973 4.5921 -65.430
##
## Step:  AIC=-74.21
## l_rate ~ slim + htype + l_len + l_adt + l_trks + l_sigs1
##
##           Df Sum of Sq    RSS     AIC
## - l_trks   1   0.14288 3.8097 -74.714
## <none>                 3.6668 -74.205
## - l_adt    1   0.31064 3.9775 -73.034
## - l_len    1   0.94373 4.6106 -67.273
## - htype    3   1.51292 5.1797 -66.733
## - l_sigs1  1   1.15981 4.8266 -65.487
## - slim     1   1.20713 4.8740 -65.107
##
## Step:  AIC=-74.71
## l_rate ~ slim + htype + l_len + l_adt + l_sigs1
##
##           Df Sum of Sq    RSS     AIC
## <none>                 3.8097 -74.714
## - l_adt    1   0.28821 4.0979 -73.870
## - htype    3   1.68565 5.4954 -66.427
## - slim     1   1.15948 4.9692 -66.352
## - l_len    1   1.24891 5.0586 -65.656
## - l_sigs1  1   1.56367 5.3734 -63.302
```

```r
summary(tight_mod)
```

```
##
## Call:
## lm(formula = l_rate ~ slim + htype + l_len + l_adt + l_sigs1,
##     data = l_dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81273 -0.17834 -0.03031  0.13832  0.66173
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.45541    0.98737   6.538 2.68e-07 ***
## slim        -0.04290    0.01397  -3.072  0.00441 **
## htypeMA     -0.38446    0.36526  -1.053  0.30067
## htypeMC     -0.17862    0.48529  -0.368  0.71533
## htypePA     -0.71475    0.28662  -2.494  0.01819 *
## l_len       -0.26161    0.08206  -3.188  0.00327 **
## l_adt       -0.12691    0.08287  -1.531  0.13581
## l_sigs1      0.20836    0.05841   3.567  0.00120 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3506 on 31 degrees of freedom
## Multiple R-squared:  0.7753, Adjusted R-squared:  0.7245
## F-statistic: 15.28 on 7 and 31 DF,  p-value: 1.835e-08
```

```r
# Let's get our coefficients, R2, and SE.
round(tight_mod$coefficients,3)
```

```
## (Intercept)        slim     htypeMA     htypeMC     htypePA       l_len
##       6.455      -0.043      -0.384      -0.179      -0.715      -0.262
##       l_adt     l_sigs1
##      -0.127       0.208
```

```r
round(summary(tight_mod)$adj.r.squared,3)
```

```
## [1] 0.725
```

```r
round(summary(tight_mod)$sigma,3)
```

```
## [1] 0.351
```

Prediction equation for tight model:
Log2(rate) = 6.455 - (0.262 * Log2(len)) - (0.127 * Log2(adt)) + (0.208 * Log2(sigs1)) - (0.043 * slim) - (0.384 * htype[MA]) - (0.179 * htype[MC]) - (0.715 * htype[PA])

Significant Predictors:
Log2(len) [p_val = 0.0033]: Log2 of the length of road segment in miles
Log2(sigs1) [p_val = 0.0012]: Log2 of number of signals per mile of roadway
slim [p_val = 0.0044]: Speed limit of road in 1973
htype[PA] [p_val = 0.0182]: Type of road = PA

The R2 of this tight model is 0.725 with a p_value of 1.835e-08. Both the R2 and the p_value are better for the tight model than the full model. The full model has an R2 of 0.683 which is a bit lower than the R2 for the tight model, indicating the tight model provides a better fit, and is a bit more significant since it as p_value of 1.835e-08 whereas the full model has a larger p-value of 1.247e-05.

The standard deviation of the error, SE, for the tight model is 0.351 whereas the SE for the full model is the larger and thus worse 0.376.

5. What is the total number of all possible regressions for the data on hand?
   Use the R function 'regsubsets' (package = "leaps") on the highway data.
   Explain the output.

The # of predictors we have here is 11, but 1 is categorical with 4 levels, which will be broken into 3 dummy variables (1 held as baseline). These 3 dummy variables replace the one htype variable so in total we have 13 predictors. So the number of all possible regressions we can do for Highway1 is 8191.

```r
# Number of regressions possible.
(2^13) - 1
```

```
## [1] 8191
```

```r
# Let's search for the best model
sub_mod <- regsubsets(l_rate ~ ., data = l_dat)
summary(sub_mod)
```

```
## Subset selection object
## Call: regsubsets.formula(l_rate ~ ., data = l_dat)
## 13 Variables  (and intercept)
##         Forced in Forced out
## slim        FALSE      FALSE
## shld        FALSE      FALSE
## lane        FALSE      FALSE
## acpt        FALSE      FALSE
## itg         FALSE      FALSE
## lwid        FALSE      FALSE
## htypeMA     FALSE      FALSE
## htypeMC     FALSE      FALSE
## htypePA     FALSE      FALSE
## l_len       FALSE      FALSE
## l_adt       FALSE      FALSE
## l_trks      FALSE      FALSE
## l_sigs1     FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##          slim shld lane acpt itg lwid htypeMA htypeMC htypePA l_len l_adt
## 1  ( 1 ) "*"  " "  " "  " "  " " " "  " "     " "     " "     " "   " "
## 2  ( 1 ) "*"  " "  " "  " "  " " " "  " "     " "     " "     "*"   " "
## 3  ( 1 ) "*"  " "  " "  " "  " " " "  " "     " "     "*"     " "   " "
## 4  ( 1 ) "*"  " "  " "  " "  " " " "  " "     " "     "*"     "*"   " "
## 5  ( 1 ) "*"  " "  " "  " "  " " " "  " "     " "     "*"     "*"   "*"
## 6  ( 1 ) "*"  " "  " "  " "  " " " "  "*"     " "     "*"     "*"   "*"
## 7  ( 1 ) "*"  " "  " "  " "  " " " "  "*"     " "     "*"     "*"   "*"
## 8  ( 1 ) "*"  " "  " "  "*"  " " " "  "*"     " "     "*"     "*"   "*"
##          l_trks l_sigs1
## 1  ( 1 ) " "    " "
## 2  ( 1 ) " "    " "
## 3  ( 1 ) " "    "*"
## 4  ( 1 ) " "    "*"
## 5  ( 1 ) " "    "*"
## 6  ( 1 ) " "    "*"
## 7  ( 1 ) "*"    "*"
## 8  ( 1 ) "*"    "*"
```

We can see that we have 13 variables, which is what we expected and thus there would be 8191 total possible regression models. However, the regsubsets function in the leaps package has a default limit of 8 predictors, so it stops after generating the best 8 predictor model, meaning it did a max of only 2^8 - 1 regressions. You are able to change this limitation according to the documentation, but I did not do this here. The output shows the predictors used in the top models from a regression model using only 1 predictor up to a model using 8 (the max) predictors. We can see that the best 1 predictor model uses 'slim', and at the 8 predictor model, it uses just 'slim', 'htype[MA]', 'htype[PA]', 'log2(len)', 'log2(adt)', 'log2(trks)', 'log2(sigs1)'.

6. Apply the 'forward selection procedure' on the 'highway1' data.
   Write the final prediction model.
   Compare and contrast this model with the one in Question 4.

```r
# First create the null model
null_mod <- lm(l_rate ~ 1, data = l_dat)

# Same as in Q4 but using forward instead of backward.
fwd_mod <- stepAIC(null_mod, direction = "forward",
                   scope = list(lower = null_mod, upper = mod_full))
```

```
## Start:  AIC=-30.5
## l_rate ~ 1
##
##           Df Sum of Sq     RSS     AIC
## + slim     1    8.0771  8.8740 -53.737
## + acpt     1    7.4345  9.5166 -51.011
## + l_sigs1  1    6.1742 10.7768 -46.160
## + l_len    1    5.5373 11.4138 -43.921
## + l_trks   1    5.0418 11.9092 -42.264
## + shld     1    2.7536 14.1974 -35.410
## <none>                 16.9510 -30.496
## + htype    3    1.8164 15.1346 -28.916
## + lane     1    0.0138 16.9373 -28.528
## + l_adt    1    0.0131 16.9379 -28.526
## + itg      1    0.0117 16.9394 -28.523
## + lwid     1    0.0082 16.9428 -28.515
##
## Step:  AIC=-53.74
## l_rate ~ slim
##
##           Df Sum of Sq    RSS     AIC
## + l_len    1   2.76182 6.1122 -66.278
## + l_trks   1   2.00977 6.8642 -61.752
## + l_sigs1  1   1.74304 7.1309 -60.266
## + acpt     1   1.16460 7.7094 -57.224
## <none>                 8.8740 -53.737
## + lane     1   0.43269 8.4413 -53.687
## + l_adt    1   0.35790 8.5161 -53.343
## + itg      1   0.35427 8.5197 -53.326
## + shld     1   0.16994 8.7040 -52.491
## + lwid     1   0.13918 8.7348 -52.354
## + htype    3   0.36259 8.5114 -49.364
##
## Step:  AIC=-66.28
## l_rate ~ slim + l_len
##
##           Df Sum of Sq    RSS     AIC
## + acpt     1   0.60035 5.5118 -68.310
```

```
## + l_trks   1   0.54776 5.5644 -67.940
## <none>                 6.1122 -66.278
## + l_sigs1  1   0.30535 5.8068 -66.277
## + htype    3   0.70029 5.4119 -65.024
## + shld     1   0.06796 6.0442 -64.714
## + l_adt    1   0.05335 6.0588 -64.620
## + lwid     1   0.03464 6.0775 -64.500
## + lane     1   0.00714 6.1050 -64.324
## + itg      1   0.00551 6.1067 -64.313
##
## Step:  AIC=-68.31
## l_rate ~ slim + l_len + acpt
##
##           Df Sum of Sq    RSS     AIC
## + l_trks   1   0.35995 5.1519 -68.944
## <none>                 5.5118 -68.310
## + l_sigs1  1   0.24989 5.2619 -68.120
## + shld     1   0.07200 5.4398 -66.823
## + l_adt    1   0.03162 5.4802 -66.534
## + lane     1   0.03095 5.4809 -66.530
## + itg      1   0.02810 5.4837 -66.509
## + lwid     1   0.02632 5.4855 -66.497
## + htype    3   0.45265 5.0592 -65.652
##
## Step:  AIC=-68.94
## l_rate ~ slim + l_len + acpt + l_trks
##
##           Df Sum of Sq    RSS     AIC
## <none>                 5.1519 -68.944
## + shld     1   0.13591 5.0159 -67.987
## + l_sigs1  1   0.10525 5.0466 -67.749
## + l_adt    1   0.06498 5.0869 -67.439
## + htype    3   0.54012 4.6117 -67.263
## + lwid     1   0.03957 5.1123 -67.245
## + itg      1   0.02282 5.1290 -67.117
## + lane     1   0.00687 5.1450 -66.996
```

```r
summary(fwd_mod)
```

```
##
## Call:
## lm(formula = l_rate ~ slim + l_len + acpt + l_trks, data = l_dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62216 -0.25940  0.05636  0.24035  0.80295
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.011048   1.069130   5.622 2.67e-06 ***
## slim        -0.045953   0.014805  -3.104  0.00383 **
## l_len       -0.235735   0.084897  -2.777  0.00887 **
## acpt         0.015876   0.009622   1.650  0.10815
## l_trks      -0.329037   0.213484  -1.541  0.13251
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3893 on 34 degrees of freedom
## Multiple R-squared:  0.6961, Adjusted R-squared:  0.6603
## F-statistic: 19.47 on 4 and 34 DF,  p-value: 2.067e-08
```

```r
# Let's get our coefficients, R2, and SE.
round(fwd_mod$coefficients,3)
```

```
## (Intercept)        slim       l_len        acpt      l_trks
##       6.011      -0.046      -0.236       0.016      -0.329
```

```r
round(summary(fwd_mod)$adj.r.squared,3)
```

```
## [1] 0.66
```

```r
round(summary(fwd_mod)$sigma,3)
```

```
## [1] 0.389
```

The forward selection procedure produces the following prediction equation:
Log2(rate) = 6.011 - (0.236 * Log2(len)) - (0.329 * Log2(trks)) + (0.016 * acpt) - (0.046 * slim)

The forward selection procedure model produces a model using only 4 predictors whereas the backwards selection procedure model uses 7! The forward model has an R2 of 0.66 with a p-value of 2.067e-08, whereas backwards has a R2 of 0.725 with a p-vale of 1.835e-08, so the backwards model is able to fit the data a bit better and both fits have similar significance values. The forward model has a SE of 0.389 whereas the backwards model has a slightly smaller SE at 0.351. So the backwards model produces a slightly better model but with the added expense of 3 more predictors than used in the forward selection procedure generated model.