Case Study 3 - Deregulation of the Intrastate Trucking Industry

Fall 2020 - STAT 214 - Project 2

Nursima Donuk

11/20/2020

Summary:

Consider the problem of modeling the price charged for motor transport service (e.g., trucking) in Florida. In the early 1980s, several states removed regulatory constraints on the rate charged for intrastate trucking services. (Florida was the first state to embark on a deregulation policy on July 1, 1980.) Prior to this time, the state determined price schedules for motor transport service with review and approval by the Public Service Commission. Once approved, individual carriers were not allowed to deviate from these official rates. The objective of the regression analysis is twofold: (1) assess the impact of deregulation on the prices charged for motor transport service in the state of Florida, and (2) estimate a model of the supply price for predicting future prices.


```
# Install development version from GitHub
# install.packages("devtools")
# devtools::install_github("rsquaredacademy/olsrr")
library(olsrr)
library(tidyverse)
```

Getting Familiar with the Data

```
load("TRUCKING.Rdata")
head(TRUCKING)
```

```
##   PRICEPTM DISTANCE WEIGHT PCTLOAD   ORIGIN   MARKET    DEREG   CARRIER PRODUCT
## 1    19942     3.60   7.50    32.6 MIA      LARGE    YES      B             100
## 2   112162     0.25   7.50    32.6 MIA      LARGE    YES      B             100
## 3    72973     0.25  15.00    65.2 MIA      LARGE    YES      B             100
## 4    41892     0.25  24.00   100.0 MIA      LARGE    YES      B             100
## 5    23519     2.60   7.50    32.6 MIA      LARGE    YES      B             100
## 6    58221     1.50   0.25     1.1 MIA      SMALL    YES      B             100
##   LNPRICE
## 1  9.9006
## 2 11.6277
## 3 11.1978
## 4 10.6428
## 5 10.0656
## 6 10.9720
```

```
str(TRUCKING)
```

```
## 'data.frame':    134 obs. of  10 variables:
##  $ PRICEPTM: num  19942 112162 72973 41892 23519 ...
##  $ DISTANCE: num  3.6 0.25 0.25 0.25 2.6 1.5 4.8 4.8 6 3 ...
```

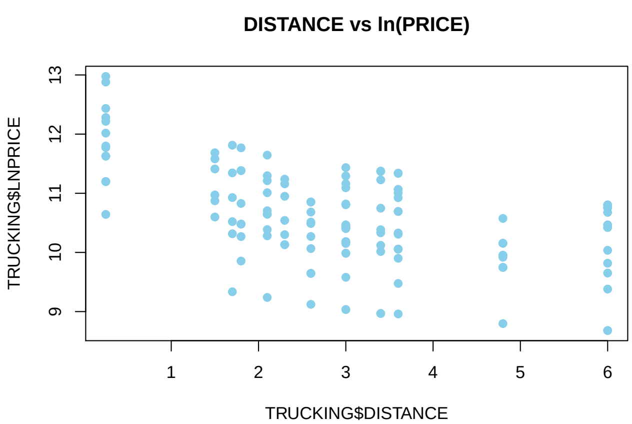
```
##  $ WEIGHT  : num  7.5 7.5 15 24 7.5 0.25 0.25 7.5 24 0.25 ...
##  $ PCTLOAD : num  32.6 32.6 65.2 100 32.6 1.1 1.1 32.6 100 1.1 ...
##  $ ORIGIN  : Factor w/ 2 levels "JAX     ","MIA     ": 2 2 2 2 2 2 2 2 2 1 ...
##  $ MARKET  : Factor w/ 2 levels "LARGE   ","SMALL   ": 1 1 1 1 1 2 1 1 2 2 ...
##  $ DEREG   : Factor w/ 2 levels "NO      ","YES     ": 2 2 2 2 2 2 2 2 2 2 ...
##  $ CARRIER : Factor w/ 1 level "B       ": 1 1 1 1 1 1 1 1 1 1 ...
##  $ PRODUCT : num  100 100 100 100 100 100 100 100 100 100 ...
##  $ LNPRICE : num  9.9 11.6 11.2 10.6 10.1 ...
```

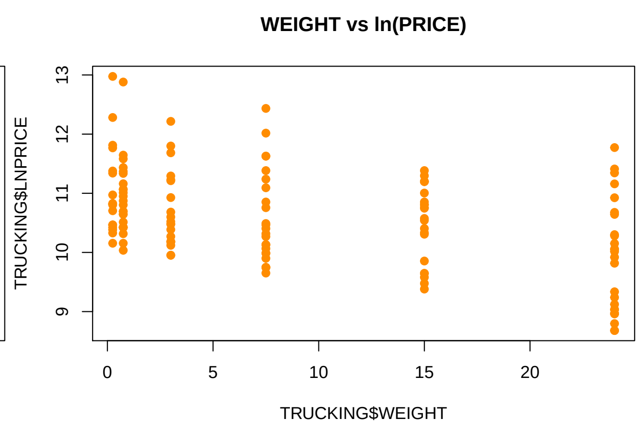
We see that the data has 10 variables and 134 observations. One of the 10 variables is `CARRIER` which is the same throughout the entire data-set. Then we have `PRICEPTM` and `LNPRICE` which are variations of the dependent variable.

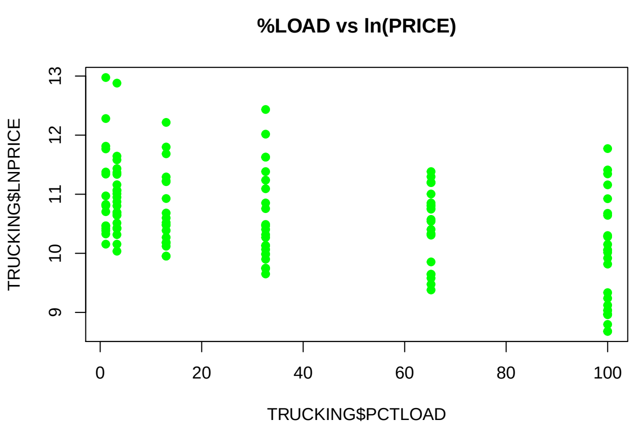
Therefore we have 6 independent variables to consider:

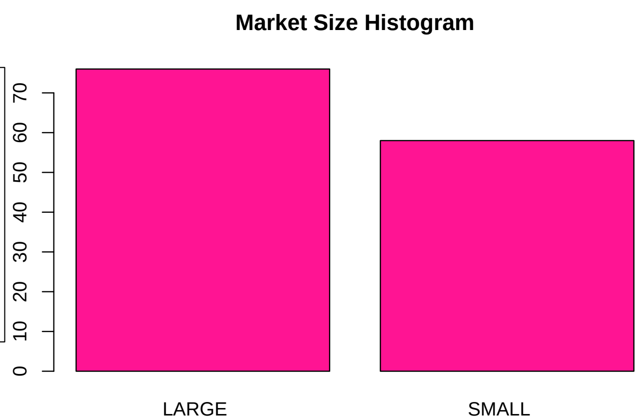
- `DISTANCE` Miles traveled (in hundreds)
- `WEIGHT` Weight of product shipped (in 1,000 pounds)
- `PCTLOAD` Percent of truck load capacity
- `ORIGIN` City of origin (JAX or MIA)
- `MARKET` Size of market destination (LARGE or SMALL)
- `DEREG` Deregulation in effect (YES or NO)

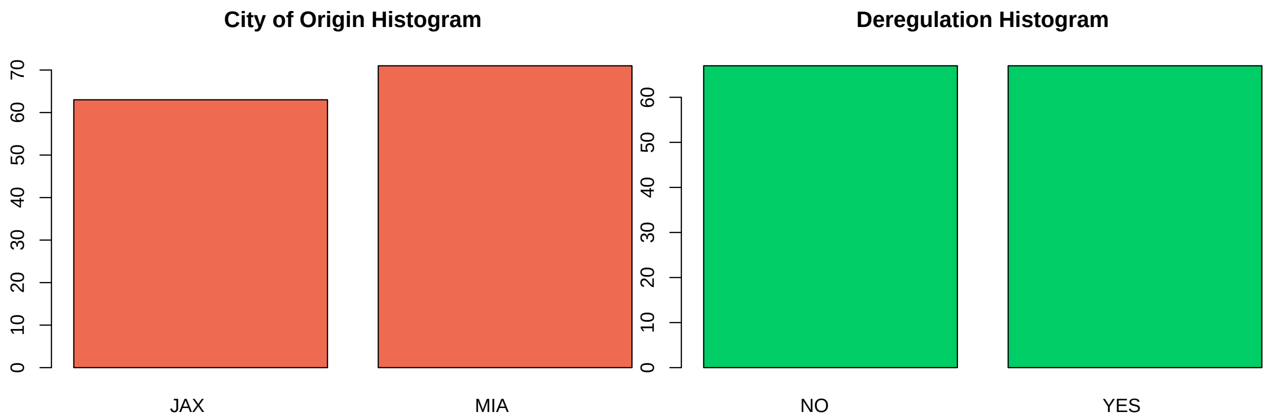
Plots











Step-wise Regression

We see that we have 6 independent variables and using all 6 of them to build a curvilinear model will require a large amount of terms, which will lead to a small degrees of freedom. So we will apply step-wise regression to choose the most relevant independent variables to the dependent variable.

```
#The plot method shows the panel of fit criteria for best subset regression methods.
model<- lm(LNPRICE ~ DISTANCE + WEIGHT + PCTLOAD + ORIGIN + MARKET + DEREG, data = TRUCKING)
k <-ols_step_both_p(model, details = T)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. DISTANCE
## 2. WEIGHT
## 3. PCTLOAD
## 4. ORIGIN
## 5. MARKET
## 6. DEREG
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - DISTANCE added
##
##                        Model Summary
## --------------------------------------------------------------
## R                      0.545       RMSE               0.692
## R-Squared              0.297       Coef. Var          6.547
## Adj. R-Squared         0.292       MSE                0.480
## Pred R-Squared         0.273       MAE                0.547
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
```

```
## ---------------------------------------------------------------------
##                Sum of
##               Squares          DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression     26.731           1         26.731    55.745    0.0000
## Residual       63.296         132          0.480
## Total          90.027         133
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## -----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## -----------------------------------------------------------------------------------------
## (Intercept)    11.424         0.128                 88.984    0.000    11.170    11.678
##    DISTANCE    -0.289         0.039       -0.545    -7.466    0.000    -0.366    -0.213
## -----------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - DEREG added
##
##                         Model Summary
## --------------------------------------------------------------
## R                       0.781       RMSE               0.518
## R-Squared               0.610       Coef. Var          4.898
## Adj. R-Squared          0.604       MSE                0.268
## Pred R-Squared          0.590       MAE                0.410
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ----------------------------------------------------------------------
##                Sum of
##               Squares          DF    Mean Square       F         Sig.
## ----------------------------------------------------------------------
## Regression     54.879           2         27.440     102.27    0.0000
## Residual       35.148         131          0.268
## Total          90.027         133
## ----------------------------------------------------------------------
##
##                                    Parameter Estimates
## -------------------------------------------------------------------------------------------
##         model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## -------------------------------------------------------------------------------------------
##   (Intercept)    11.935         0.108                 110.303    0.000    11.721    12.149
##      DISTANCE    -0.307         0.029       -0.578    -10.568    0.000    -0.364    -0.249
## DEREGYES         -0.918         0.090       -0.560    -10.243    0.000    -1.096    -0.741
## -------------------------------------------------------------------------------------------
##
##
```

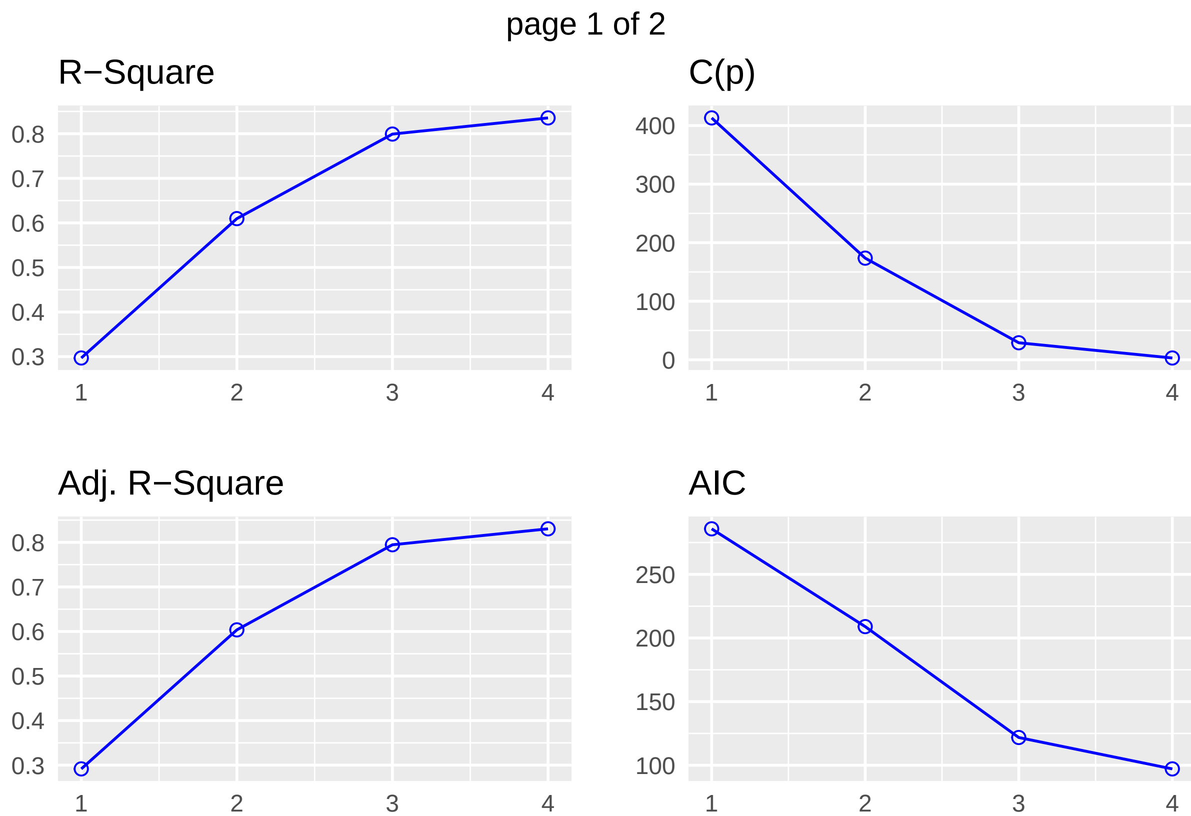
```
##
##                        Model Summary
## --------------------------------------------------------------
## R                      0.781       RMSE               0.518
## R-Squared              0.610       Coef. Var          4.898
## Adj. R-Squared         0.604       MSE                0.268
## Pred R-Squared         0.590       MAE                0.410
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ----------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square       F         Sig.
## ----------------------------------------------------------------------
## Regression     54.879          2         27.440     102.27    0.0000
## Residual       35.148        131          0.268
## Total          90.027        133
## ----------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------
##         model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
##   (Intercept)    11.935         0.108                 110.303    0.000    11.721    12.149
##      DISTANCE    -0.307         0.029       -0.578    -10.568    0.000    -0.364    -0.249
## DEREGYES         -0.918         0.090       -0.560    -10.243    0.000    -1.096    -0.741
## ----------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 3
##
## - WEIGHT added
##
##                        Model Summary
## --------------------------------------------------------------
## R                      0.894       RMSE               0.373
## R-Squared              0.799       Coef. Var          3.525
## Adj. R-Squared         0.795       MSE                0.139
## Pred R-Squared         0.784       MAE                0.282
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ----------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square       F         Sig.
## ----------------------------------------------------------------------
## Regression     71.956          3         23.985    172.541    0.0000
```

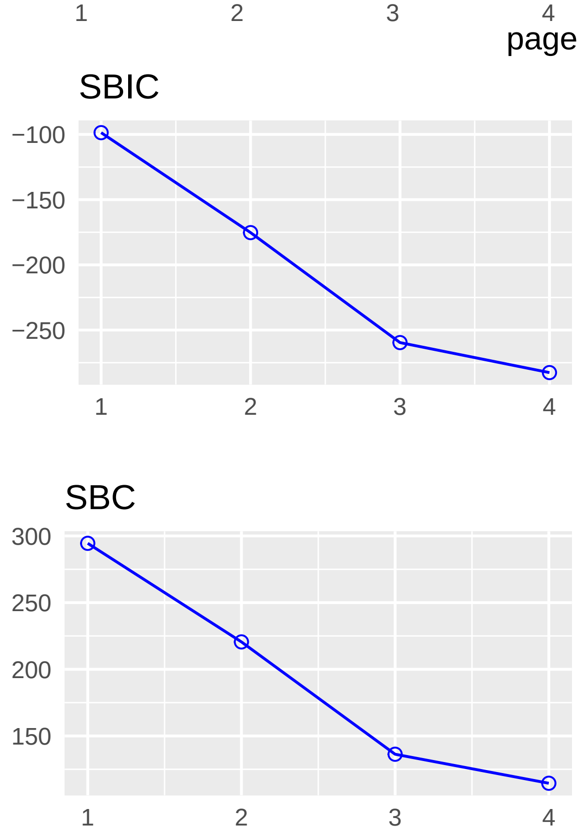
```
## Residual       18.071        130          0.139
## Total          90.027        133
## ---------------------------------------------------------------------
##
##                                     Parameter Estimates
## ------------------------------------------------------------------------------------------
##         model      Beta    Std. Error    Std. Beta        t        Sig      lower      upper
## ------------------------------------------------------------------------------------------
##   (Intercept)    12.276         0.084                  146.573    0.000     12.111     12.442
##      DISTANCE    -0.291         0.021       -0.549    -13.910    0.000     -0.333     -0.250
## DEREGYES         -0.954         0.065       -0.582    -14.765    0.000     -1.082     -0.826
##        WEIGHT    -0.041         0.004       -0.437    -11.083    0.000     -0.048     -0.033
## ------------------------------------------------------------------------------------------
##
##
##
##                         Model Summary
## -------------------------------------------------------------
## R                       0.894       RMSE                0.373
## R-Squared               0.799       Coef. Var           3.525
## Adj. R-Squared          0.795       MSE                 0.139
## Pred R-Squared          0.784       MAE                 0.282
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression     71.956          3         23.985    172.541    0.0000
## Residual       18.071        130          0.139
## Total          90.027        133
## ---------------------------------------------------------------------
##
##                                     Parameter Estimates
## ------------------------------------------------------------------------------------------
##         model      Beta    Std. Error    Std. Beta        t        Sig      lower      upper
## ------------------------------------------------------------------------------------------
##   (Intercept)    12.276         0.084                  146.573    0.000     12.111     12.442
##      DISTANCE    -0.291         0.021       -0.549    -13.910    0.000     -0.333     -0.250
## DEREGYES         -0.954         0.065       -0.582    -14.765    0.000     -1.082     -0.826
##        WEIGHT    -0.041         0.004       -0.437    -11.083    0.000     -0.048     -0.033
## ------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - ORIGIN added
##
##                         Model Summary
```

```
## -----------------------------------------------------------
## R                        0.914       RMSE              0.339
## R-Squared                0.835       Coef. Var         3.204
## Adj. R-Squared           0.830       MSE               0.115
## Pred R-Squared           0.821       MAE               0.262
## -----------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square       F         Sig.
## -------------------------------------------------------------------
## Regression     75.217          4         18.804    163.786    0.0000
## Residual       14.810        129          0.115
## Total          90.027        133
## -------------------------------------------------------------------
##
##                                    Parameter Estimates
## ---------------------------------------------------------------------------------------------
##          model      Beta    Std. Error    Std. Beta       t        Sig      lower      upper
## ---------------------------------------------------------------------------------------------
##    (Intercept)    12.157         0.079                 153.202    0.000     12.000     12.314
##       DISTANCE    -0.301         0.019       -0.567    -15.742    0.000     -0.339     -0.263
##  DEREGYES         -0.989         0.059       -0.603    -16.738    0.000     -1.106     -0.872
##         WEIGHT    -0.041         0.003       -0.440    -12.261    0.000     -0.047     -0.034
## ORIGINMIA          0.316         0.059        0.192      5.330    0.000      0.199      0.433
## ---------------------------------------------------------------------------------------------
##
##
##
##                        Model Summary
## -----------------------------------------------------------
## R                        0.914       RMSE              0.339
## R-Squared                0.835       Coef. Var         3.204
## Adj. R-Squared           0.830       MSE               0.115
## Pred R-Squared           0.821       MAE               0.262
## -----------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square       F         Sig.
## -------------------------------------------------------------------
## Regression     75.217          4         18.804    163.786    0.0000
## Residual       14.810        129          0.115
## Total          90.027        133
## -------------------------------------------------------------------
##
```

```
##                                     Parameter Estimates
## -------------------------------------------------------------------------------------------
##          model      Beta    Std. Error    Std. Beta        t        Sig      lower      upper
## -------------------------------------------------------------------------------------------
##    (Intercept)    12.157         0.079                  153.202    0.000     12.000     12.314
##       DISTANCE    -0.301         0.019       -0.567     -15.742    0.000     -0.339     -0.263
##  DEREGYES         -0.989         0.059       -0.603     -16.738    0.000     -1.106     -0.872
##         WEIGHT    -0.041         0.003       -0.440     -12.261    0.000     -0.047     -0.034
## ORIGINMIA          0.316         0.059        0.192       5.330    0.000      0.199      0.433
## -------------------------------------------------------------------------------------------
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## -------------------------------------------------------------
## R                       0.914       RMSE              0.339
## R-Squared               0.835       Coef. Var         3.204
## Adj. R-Squared          0.830       MSE               0.115
## Pred R-Squared          0.821       MAE               0.262
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## ---------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression     75.217          4         18.804    163.786    0.0000
## Residual       14.810        129          0.115
## Total          90.027        133
## ---------------------------------------------------------------------
##
##                                     Parameter Estimates
## -------------------------------------------------------------------------------------------
##          model      Beta    Std. Error    Std. Beta        t        Sig      lower      upper
## -------------------------------------------------------------------------------------------
##    (Intercept)    12.157         0.079                  153.202    0.000     12.000     12.314
##       DISTANCE    -0.301         0.019       -0.567     -15.742    0.000     -0.339     -0.263
##  DEREGYES         -0.989         0.059       -0.603     -16.738    0.000     -1.106     -0.872
##         WEIGHT    -0.041         0.003       -0.440     -12.261    0.000     -0.047     -0.034
## ORIGINMIA          0.316         0.059        0.192       5.330    0.000      0.199      0.433
## -------------------------------------------------------------------------------------------
```

```r
plot(k)
```





We can observe that the model with 4 independent variables provides us the best predictors.

- `DISTANCE` Miles traveled (in hundreds)

- `WEIGHT` Weight of product shipped (in 1,000 pounds)
- `ORIGIN` City of origin (JAX or MIA)
- `DEREG` Deregulation in effect (YES or NO)

Distance and weight are quantitative variables while origin and deregulation is qualitative.

$$x_1 = \text{Distance shipped}$$

$$x_2 = \text{Weight of product}$$

$$x_3 = \begin{cases} 1 & \text{if deregulation in effect} \\ 0 & \text{if not} \end{cases}$$

$$x_4 = \begin{cases} 1 & \text{if originate in Miami} \\ 0 & \text{if originate in Jacksonville} \end{cases}$$

Building the First Models

Renaming variables

```
Y <- TRUCKING$LNPRICE
X1 <- TRUCKING$DISTANCE
X2 <- TRUCKING$WEIGHT
X3 <- TRUCKING$DEREG
X4 <- TRUCKING$ORIGIN
```

The Complete Second Order Model

```
model1 <- lm(Y ~ X1 + X2 + X1*X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3*X4 + X1*X3 + X1*X4 + X1*X3*X4 + X2*
summary(model1)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 +
##     X4 + X3 * X4 + X1 * X3 + X1 * X4 + X1 * X3 * X4 + X2 * X3 +
##     X2 * X4 + X2 * X3 * X4 + X1 * X2 * X3 + X1 * X2 * X4 + X1 *
##     X2 * X3 * X4 + I(X1^2) * X3 + I(X1^2) * X4 + I(X1^2) * X3 *
##     X4 + I(X2^2) * X3 + I(X2^2) * X4 + I(X2^2) * X3 * X4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.56178 -0.11493  0.01701  0.13499  0.53106
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   12.5158767  0.9544115  13.114  < 2e-16 ***
## X1                            -0.8991875  0.7340996  -1.225  0.22324
## X2                             0.0242073  0.0288603   0.839  0.40341
## I(X1^2)                        0.1514393  0.1345553   1.125  0.26284
## I(X2^2)                       -0.0001020  0.0007696  -0.133  0.89479
## X3YES                         -1.1264341  1.4910477  -0.755  0.45158
## X4MIA                          0.2761451  0.9633225   0.287  0.77491
## X1:X2                         -0.0207090  0.0067322  -3.076  0.00265 **
## X3YES     :X4MIA               0.4969434  1.5029026   0.331  0.74153
```

```
## X1:X3YES                          0.4819536  1.1588221   0.416  0.67829
## X1:X4MIA                          0.0695520  0.7388206   0.094  0.92517
## X2:X3YES                         -0.0948579  0.0447667  -2.119  0.03635 *
## X2:X4MIA                         -0.0526100  0.0352796  -1.491  0.13876
## I(X1^2):X3YES                    -0.1167119  0.2191804  -0.532  0.59546
## I(X1^2):X4MIA                    -0.0727509  0.1350954  -0.539  0.59131
## I(X2^2):X3YES                     0.0004377  0.0011854   0.369  0.71262
## I(X2^2):X4MIA                     0.0001108  0.0010677   0.104  0.91751
## X1:X3YES     :X4MIA              -0.5402672  1.1644024  -0.464  0.64357
## X2:X3YES     :X4MIA               0.0682541  0.0522034   1.307  0.19378
## X1:X2:X3YES                       0.0220705  0.0107754   2.048  0.04292 *
## X1:X2:X4MIA                       0.0235483  0.0070932   3.320  0.00122 **
## I(X1^2):X3YES     :X4MIA          0.1342074  0.2198421   0.610  0.54281
## I(X2^2):X3YES     :X4MIA         -0.0002761  0.0015707  -0.176  0.86079
## X1:X2:X3YES     :X4MIA           -0.0269349  0.0112668  -2.391  0.01852 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2358 on 110 degrees of freedom
## Multiple R-squared:  0.932,  Adjusted R-squared:  0.9178
## F-statistic: 65.59 on 23 and 110 DF,  p-value: < 2.2e-16
```

Taking out Squared Terms

We will create another model, that is the same as the above except that it does not have the squared terms.

```r
model2 <- lm(Y ~ X1 + X2 + X1*X2 +X3 + X4 + X3*X4 + X1*X3 + X1*X4 + X1*X3*X4 + X2*X3 + X2*X4 + X2*X3*X4
summary(model2)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X1 * X2 + X3 + X4 + X3 * X4 + X1 *
##     X3 + X1 * X4 + X1 * X3 * X4 + X2 * X3 + X2 * X4 + X2 * X3 *
##     X4 + X1 * X2 * X3 + X1 * X2 * X4 + X1 * X2 * X3 * X4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.84999 -0.18273  0.01177  0.17947  0.77398
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                11.472100   0.348461  32.922   <2e-16 ***
## X1                         -0.077415   0.118534  -0.653   0.5150
## X2                          0.021123   0.025459   0.830   0.4084
## X3YES                      -0.331415   0.555697  -0.596   0.5521
## X4MIA                       0.882581   0.376595   2.344   0.0208 *
## X1:X2                      -0.020639   0.008957  -2.304   0.0230 *
## X3YES     :X4MIA           -0.424913   0.592346  -0.717   0.4746
## X1:X3YES                   -0.157470   0.196829  -0.800   0.4253
## X1:X4MIA                   -0.237177   0.124228  -1.909   0.0587 .
## X2:X3YES                   -0.082562   0.041064  -2.011   0.0467 *
## X2:X4MIA                   -0.056221   0.028362  -1.982   0.0498 *
## X1:X3YES     :X4MIA         0.152258   0.204769   0.744   0.4586
## X2:X3YES     :X4MIA         0.071941   0.044639   1.612   0.1097
## X1:X2:X3YES                 0.021777   0.014480   1.504   0.1353
## X1:X2:X4MIA                 0.023977   0.009459   2.535   0.0126 *
```

```
## X1:X2:X3YES     :X4MIA      -0.025949   0.015136  -1.714   0.0891 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3212 on 118 degrees of freedom
## Multiple R-squared:  0.8647, Adjusted R-squared:  0.8476
## F-statistic:  50.3 on 15 and 118 DF,  p-value: < 2.2e-16
```

Keep the Squared Terms and Remove Interaction Between Qualitative and Quantitative

```
model3 <- lm(Y ~ X1 + X2 + X1*X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3*X4)
summary(model3)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 +
##     X4 + X3 * X4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.71405 -0.17870  0.03056  0.20084  0.83521
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           12.8751219  0.1274209 101.044  < 2e-16 ***
## X1                    -0.7746834  0.0642515 -12.057  < 2e-16 ***
## X2                    -0.0457822  0.0120903  -3.787 0.000236 ***
## I(X1^2)                0.0751044  0.0096511   7.782 2.32e-12 ***
## I(X2^2)                0.0001066  0.0004468   0.239 0.811865
## X3YES                 -0.9776064  0.0717005 -13.635  < 2e-16 ***
## X4MIA                  0.0226420  0.0798112   0.284 0.777114
## X1:X2                  0.0005142  0.0017966   0.286 0.775208
## X3YES     :X4MIA       0.0308881  0.0987742   0.313 0.755019
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2822 on 125 degrees of freedom
## Multiple R-squared:  0.8894, Adjusted R-squared:  0.8823
## F-statistic: 125.7 on 8 and 125 DF,  p-value: < 2.2e-16
```

Only Remove Squared Interactions

```
model4 <- lm(Y ~ X1 + X2 + X1*X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3*X4 + X1*X3 + X1*X4 + X1*X3*X4 + X2*
summary(model4)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 +
##     X4 + X3 * X4 + X1 * X3 + X1 * X4 + X1 * X3 * X4 + X2 * X3 +
##     X2 * X4 + X2 * X3 * X4 + X1 * X2 * X3 + X1 * X2 * X4 + X1 *
##     X2 * X3 * X4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59556 -0.10610  0.01649  0.11986  0.55701
##
```

```
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   1.208e+01  2.587e-01  46.712  < 2e-16 ***
## X1                           -5.530e-01  9.648e-02  -5.731 8.02e-08 ***
## X2                            1.889e-02  2.120e-02   0.891 0.374766
## I(X1^2)                       8.738e-02  8.274e-03  10.560  < 2e-16 ***
## I(X2^2)                       8.196e-05  3.735e-04   0.219 0.826705
## X3YES                        -3.852e-01  4.001e-01  -0.963 0.337705
## X4MIA                         7.604e-01  2.714e-01   2.802 0.005952 **
## X1:X2                        -2.041e-02  6.487e-03  -3.146 0.002101 **
## X3YES     :X4MIA             -3.529e-01  4.266e-01  -0.827 0.409765
## X1:X3YES                     -1.316e-01  1.417e-01  -0.929 0.355058
## X1:X4MIA                     -3.337e-01  8.995e-02  -3.710 0.000320 ***
## X2:X3YES                     -8.258e-02  2.956e-02  -2.793 0.006104 **
## X2:X4MIA                     -4.830e-02  2.049e-02  -2.357 0.020096 *
## X1:X3YES     :X4MIA           1.821e-01  1.475e-01   1.235 0.219378
## X2:X3YES     :X4MIA           5.936e-02  3.217e-02   1.845 0.067524 .
## X1:X2:X3YES                   2.136e-02  1.043e-02   2.048 0.042807 *
## X1:X2:X4MIA                   2.319e-02  6.846e-03   3.388 0.000963 ***
## X1:X2:X3YES     :X4MIA       -2.600e-02  1.090e-02  -2.386 0.018628 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2312 on 116 degrees of freedom
## Multiple R-squared:  0.9311, Adjusted R-squared:  0.921
## F-statistic: 92.21 on 17 and 116 DF,  p-value: < 2.2e-16
```

Choosing a model at $(\alpha = 0.01)$

We can observe that all models resulted in a small p-value from the global F-test. Meaning they all are statistically useful for predicting trucking price. We will one by one compare the models.

Model 1 vs Model 2: We can observe that Model 1 has a higher adjusted R squared and more statistically significant terms. We will conduct a partial F-test to see if the full model is statistically a better better predictor than the reduced model (Model 2).

$$H_0 : \beta_4 = \beta_5 = \beta_{18} = \beta_{19} = \beta_{20} = \beta_{21} = \beta_{22} = \beta_{23} = 0$$

$$H_a : \text{At least one of the quadratic } \beta\text{'s in Model 1 differs from 0}$$

```r
anova(model1, model2)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3 * X4 +
##     X1 * X3 + X1 * X4 + X1 * X3 * X4 + X2 * X3 + X2 * X4 + X2 *
##     X3 * X4 + X1 * X2 * X3 + X1 * X2 * X4 + X1 * X2 * X3 * X4 +
##     I(X1^2) * X3 + I(X1^2) * X4 + I(X1^2) * X3 * X4 + I(X2^2) *
##     X3 + I(X2^2) * X4 + I(X2^2) * X3 * X4
## Model 2: Y ~ X1 + X2 + X1 * X2 + X3 + X4 + X3 * X4 + X1 * X3 + X1 * X4 +
##     X1 * X3 * X4 + X2 * X3 + X2 * X4 + X2 * X3 * X4 + X1 * X2 *
##     X3 + X1 * X2 * X4 + X1 * X2 * X3 * X4
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1    110  6.1186
## 2    118 12.1764 -8   -6.0578 13.613 1.465e-13 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The small p-value suggests that we can reject the null hypothesis. Concluding that the quadratic terms in Model 1 are statistically significant.

Model 1 vs Model 3: We can see that the adjusted R squared of Model 3 is even lower than Model 2's. We will conduct a partial F-test to see if the reduced model is statistically worse than Model 1 (complete second order model).

$$H_0 : \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = \beta_{15} = \beta_{16} = \beta_{17} = \beta_{18} = \beta_{19} = \beta_{20} = \beta_{21} = \beta_{22} = \beta_{23} = 0$$

$$H_a : \text{At least one of the QNxQL interaction } \beta\text{'s in Model 1 differs from 0}$$

```
anova(model1, model3)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3 * X4 +
##     X1 * X3 + X1 * X4 + X1 * X3 * X4 + X2 * X3 + X2 * X4 + X2 *
##     X3 * X4 + X1 * X2 * X3 + X1 * X2 * X4 + X1 * X2 * X3 * X4 +
##     I(X1^2) * X3 + I(X1^2) * X4 + I(X1^2) * X3 * X4 + I(X2^2) *
##     X3 + I(X2^2) * X4 + I(X2^2) * X3 * X4
## Model 2: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3 * X4
##   Res.Df    RSS  Df Sum of Sq      F    Pr(>F)
## 1    110 6.1186
## 2    125 9.9548 -15   -3.8361 4.5977 9.838e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The small p-value suggests that the QNxQL interation terms are significant and at least one of the coefficients differ from 0. Therefore we will continue with Model 1.

Model 1 vs Model 4: We can observe that Model 4 has a significant amount of more statistically significant terms. Also Model 4 has a higher adjusted R squared value. Conducting a partial F-test will show which model is statistically significant aside from the observations.

$$H_0 : \beta_{18} = \beta_{19} = \beta_{20} = \beta_{21} = \beta_{22} = \beta_{23} = 0$$

$$H_a : \text{At least one of the qualitative-quadratic interaction } \beta\text{'s in Model 1 differs from 0}$$

```
anova(model1, model4)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3 * X4 +
##     X1 * X3 + X1 * X4 + X1 * X3 * X4 + X2 * X3 + X2 * X4 + X2 *
##     X3 * X4 + X1 * X2 * X3 + X1 * X2 * X4 + X1 * X2 * X3 * X4 +
##     I(X1^2) * X3 + I(X1^2) * X4 + I(X1^2) * X3 * X4 + I(X2^2) *
##     X3 + I(X2^2) * X4 + I(X2^2) * X3 * X4
## Model 2: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3 * X4 +
##     X1 * X3 + X1 * X4 + X1 * X3 * X4 + X2 * X3 + X2 * X4 + X2 *
##     X3 * X4 + X1 * X2 * X3 + X1 * X2 * X4 + X1 * X2 * X3 * X4
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    110 6.1186
## 2    116 6.2030 -6 -0.084382 0.2528 0.9572
```

The large p-value suggests that we fail to reject the null hypothesis and therefore we choose Model 4 to be the better predictor.

Building More Models

Drop Terms Containing X4 from Model 4

```
model5 <- lm(Y ~ X1 + X2 + X1*X2 + I(X1^2) + I(X2^2) + X3 + X1*X3 + X2*X3 + X1*X2*X3)
summary(model5)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 +
##     X1 * X3 + X2 * X3 + X1 * X2 * X3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6398 -0.1766  0.0191  0.1435  0.6524
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      12.8495047  0.1113510 115.396  < 2e-16 ***
## X1               -0.8347011  0.0534069 -15.629  < 2e-16 ***
## X2               -0.0380653  0.0121254  -3.139  0.00212 **
## I(X1^2)           0.0830077  0.0070383  11.794  < 2e-16 ***
## I(X2^2)           0.0001873  0.0003943   0.475  0.63552
## X3YES            -0.7866917  0.1366532  -5.757 6.34e-08 ***
## X1:X2             0.0018214  0.0021559   0.845  0.39982
## X1:X3YES          0.0323006  0.0421641   0.766  0.44509
## X2:X3YES         -0.0190077  0.0110158  -1.725  0.08693 .
## X1:X2:X3YES      -0.0033701  0.0032005  -1.053  0.29439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2495 on 124 degrees of freedom
## Multiple R-squared:  0.9142, Adjusted R-squared:  0.908
## F-statistic: 146.9 on 9 and 124 DF,  p-value: < 2.2e-16
```

Drop Terms Containing X3 from Model 4

```
model6 <- lm(Y ~ X1 + X2 + X1*X2 + I(X1^2) + I(X2^2) + X4 + X1*X4 + X2*X4 + X1*X2*X4)
summary(model6)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X4 +
##     X1 * X4 + X2 * X4 + X1 * X2 * X4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.10610 -0.41914 -0.09361  0.45041  1.32088
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      11.5555587  0.4866063  23.747  < 2e-16 ***
## X1               -0.5037576  0.1922962  -2.620   0.0099 **
```

```
## X2                 0.0081068  0.0424667   0.191   0.8489
## I(X1^2)            0.0922534  0.0191887   4.808 4.34e-06 ***
## I(X2^2)            0.0001259  0.0008889   0.142   0.8876
## X4MIA              0.8780703  0.4966336   1.768   0.0795 .
## X1:X2             -0.0194064  0.0122038  -1.590   0.1143
## X1:X4MIA          -0.3836006  0.1708012  -2.246   0.0265 *
## X2:X4MIA          -0.0515772  0.0378926  -1.361   0.1759
## X1:X2:X4MIA        0.0210225  0.0128192   1.640   0.1036
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.557 on 124 degrees of freedom
## Multiple R-squared:  0.5726, Adjusted R-squared:  0.5416
## F-statistic: 18.46 on 9 and 124 DF,  p-value: < 2.2e-16
```

Drop all Qualitative-Qualitative Interactions

```r
model7 <- lm(Y ~ X1 + X2 + X1*X2 + I(X1^2) + I(X2^2) + X3 + X4 + X1*X3 + X1*X4+ X2*X3 + X2*X4 + X1*X2*X3
summary(model7)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 +
##     X4 + X1 * X3 + X1 * X4 + X2 * X3 + X2 * X4 + X1 * X2 * X3 +
##     X1 * X2 * X4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.52773 -0.14083 -0.00927  0.14134  0.58708
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       12.1914422  0.2158264  56.487  < 2e-16 ***
## X1                -0.5979770  0.0842487  -7.098 9.64e-11 ***
## X2                -0.0059761  0.0185722  -0.322 0.748182
## I(X1^2)            0.0857500  0.0083444  10.276  < 2e-16 ***
## I(X2^2)            0.0001420  0.0003773   0.376 0.707262
## X3YES             -0.7818945  0.1290013  -6.061 1.61e-08 ***
## X4MIA              0.6768020  0.2103468   3.218 0.001663 **
## X1:X2             -0.0107837  0.0053020  -2.034 0.044168 *
## X1:X3YES           0.0399004  0.0399904   0.998 0.320409
## X1:X4MIA          -0.2746418  0.0726651  -3.780 0.000246 ***
## X2:X3YES          -0.0209445  0.0104470  -2.005 0.047232 *
## X2:X4MIA          -0.0261968  0.0160964  -1.627 0.106256
## X1:X2:X3YES       -0.0033175  0.0030303  -1.095 0.275811
## X1:X2:X4MIA        0.0129784  0.0054379   2.387 0.018565 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2355 on 120 degrees of freedom
## Multiple R-squared:  0.926,  Adjusted R-squared:  0.918
## F-statistic: 115.6 on 13 and 120 DF,  p-value: < 2.2e-16
```

Choosing the Final Model at ($\alpha = 0.01$)

Model 4 vs Model 5: We can observe that the adjusted R squared in both models is high. Conducting a partial F-test to compare these models results in the following output:

$$H_0 : \beta_7 = \beta_8 = \beta_{10} = \beta_{11} = \beta_{13} = \beta_{14} = \beta_{16} = \beta_{17} = 0$$

$$H_a : \text{At least one of the origin } \beta\text{'s in Model 4 differs from 0}$$

```r
anova(model4, model5)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3 * X4 +
##     X1 * X3 + X1 * X4 + X1 * X3 * X4 + X2 * X3 + X2 * X4 + X2 *
##     X3 * X4 + X1 * X2 * X3 + X1 * X2 * X4 + X1 * X2 * X3 * X4
## Model 2: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 + X1 * X3 + X2 *
##     X3 + X1 * X2 * X3
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1    116 6.203
## 2    124 7.722 -8    -1.519 3.5507 0.00103 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The small p-value ($< .01$) means that we reject the null hypothesis and we conclude that origin terms are statistically significant. Therefore, we continue testing with Model 4.

Model 4 vs Model 6: It can be observed that the adjusted R squared value for Model 6 is significantly lower than Model 4's. We can conduct a partial F-test to find out if Model 4 is statistically a better predictor than Model 6.

$$H_0 : \beta_6 = \beta_8 = \beta_9 = \beta_{11} = \beta_{12} = \beta_{14} = \beta_{15} = \beta_{17} = 0$$

$$H_a : \text{At least one of the deregulation } \beta\text{'s in Model 4 differs from 0}$$

```r
anova(model4, model6)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3 * X4 +
##     X1 * X3 + X1 * X4 + X1 * X3 * X4 + X2 * X3 + X2 * X4 + X2 *
##     X3 * X4 + X1 * X2 * X3 + X1 * X2 * X4 + X1 * X2 * X3 * X4
## Model 2: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X4 + X1 * X4 + X2 *
##     X4 + X1 * X2 * X4
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    116  6.203
## 2    124 38.476 -8   -32.273 75.441 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The small p-value resulting from this test leads us to reject the null hypothesis. Concluding that Model 4 is a statistically better predictor for trucking price.

Model 4 vs Model 7: We can see that the adjusted R squared values for these models do not differ significantly. Conducting a partial F-test results in:

$$H_0 = \beta_8 = \beta_{11} = \beta_{14} = \beta_{17} = 0$$

$$H_a : \text{At least one of the QLxQL interaction } \beta\text{'s in Model 4 differs from 0}$$

```r
anova(model4, model7)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 + X4 + X3 * X4 +
##     X1 * X3 + X1 * X4 + X1 * X3 * X4 + X2 * X3 + X2 * X4 + X2 *
##     X3 * X4 + X1 * X2 * X3 + X1 * X2 * X4 + X1 * X2 * X3 * X4
## Model 2: Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 + X4 + X1 * X3 +
##     X1 * X4 + X2 * X3 + X2 * X4 + X1 * X2 * X3 + X1 * X2 * X4
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    116 6.2030
## 2    120 6.6577 -4  -0.45465 2.1256  0.082 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It can be observed that the p-value is $> .01$, meaning we fail to reject the null hypothesis that all QLxQL interaction terms are 0. Leading us to choose Model 7 as our final model.

Impact of Deregulation

Now that we have chosen our model, let us observe the impact of deregulation on trucking price.

```r
summary(model7)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 +
##     X4 + X1 * X3 + X1 * X4 + X2 * X3 + X2 * X4 + X1 * X2 * X3 +
##     X1 * X2 * X4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.52773 -0.14083 -0.00927  0.14134  0.58708
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      12.1914422  0.2158264  56.487  < 2e-16 ***
## X1               -0.5979770  0.0842487  -7.098 9.64e-11 ***
## X2               -0.0059761  0.0185722  -0.322 0.748182
## I(X1^2)           0.0857500  0.0083444  10.276  < 2e-16 ***
## I(X2^2)           0.0001420  0.0003773   0.376 0.707262
## X3YES            -0.7818945  0.1290013  -6.061 1.61e-08 ***
## X4MIA             0.6768020  0.2103468   3.218 0.001663 **
## X1:X2            -0.0107837  0.0053020  -2.034 0.044168 *
## X1:X3YES          0.0399004  0.0399904   0.998 0.320409
## X1:X4MIA         -0.2746418  0.0726651  -3.780 0.000246 ***
## X2:X3YES         -0.0209445  0.0104470  -2.005 0.047232 *
## X2:X4MIA         -0.0261968  0.0160964  -1.627 0.106256
## X1:X2:X3YES      -0.0033175  0.0030303  -1.095 0.275811
## X1:X2:X4MIA       0.0129784  0.0054379   2.387 0.018565 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2355 on 120 degrees of freedom
```

```
## Multiple R-squared:  0.926,  Adjusted R-squared:  0.918
## F-statistic: 115.6 on 13 and 120 DF,  p-value: < 2.2e-16
```

$$\hat{y} = 12.192 - .598x_1 - .00598x_2 - .01078x_1x_2 + .086x_1^2 + .00014x_2^2$$

$$+.677x_4 - .275x_1x_4 - .026x_2x_4 + .013x_1x_2x_4 - .782x_3$$

$$+.0399x_1x_3 - .021x_2x_3 - .0033x_1x_2x_3$$

A good way to assess the impact of deregulation is to hold all but one independent variable fixed. Suppose weight of shipment is 15,000 pounds and consider only shipments originating from Jacksonville ($x_2 = 15$ and $x_4 = 0$). Substituting these values into the prediction equation results in:

$$\hat{y} = 12.192 - .598x_1 - .00598(15) - .01078x_1(15) + .086x_1^2 + .00014(15)^2$$

$$+.677(0) - .275x_1(0) - .026(15)(0) + .013x_1(15)(0) - .782x_3$$

$$+.0399x_1x_3 - .021(15)x_3 - .0033x_1(15)x_3$$

$$= 12.192 - .760x_1 + .086x_1^2 - 1.097x_3 - .0096x_1x_3$$

To see the impact of deregulation now we will plug in $x_3 = 1$ (deregulated) and $x_3 = 0$ (regulated), shown below:

$$\text{Regulated}(x_3 = 0) : \hat{y} = 12.192 - .760x_1 + .086x_1^2 - 1.097(0) - .0096x_1(0)$$

$$= 12.192 - .760x_1 + .086x_1^2$$

$$\text{Deregulation}(x_3 = 1) : \hat{y} = 12.192 - .760x_1 + .086x_1^2 - 1.097(1) - .0096x_1(1)$$

$$= 11.037 - .7696x_1 + .086x_1^2$$

We can see that the y-intercept for the regulated prices is larger than the y-intercept for the deregulated prices. The equations have the same curvature but the shift parameter differs.

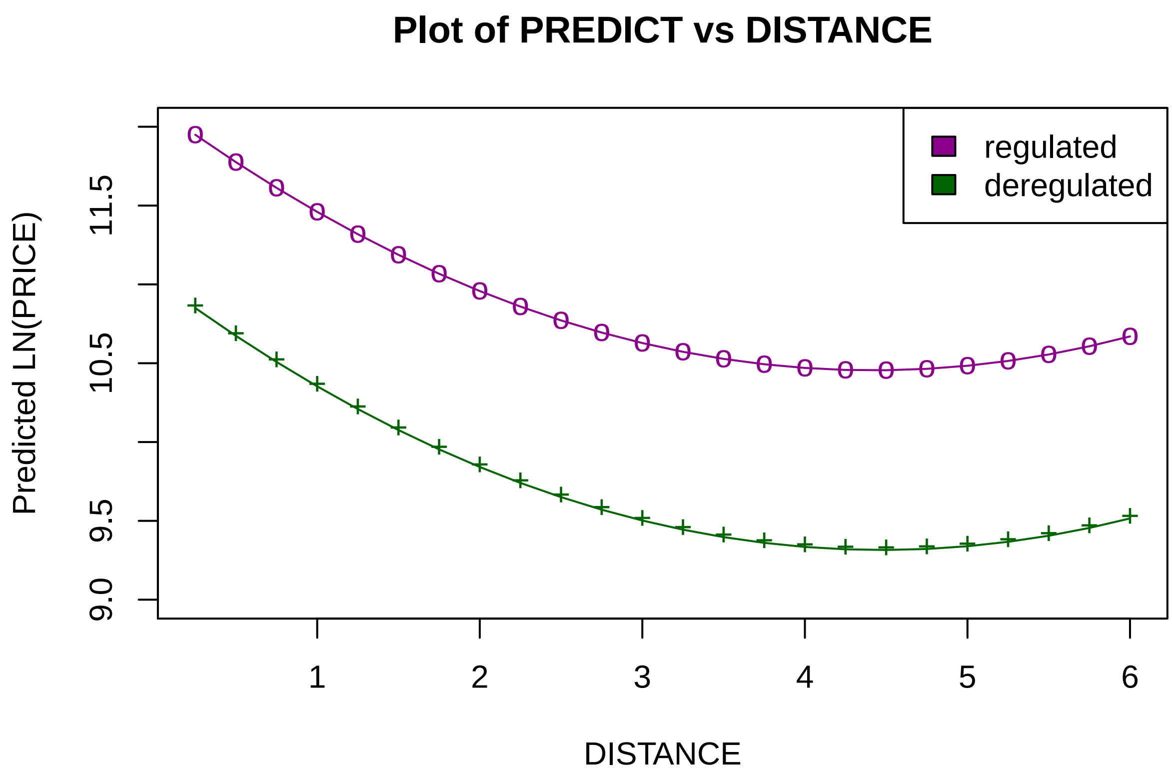
```r
reg <- function(x) {
  yint<- 12.134
  shift <- 0.76*x
  curve <- 0.086*x*x
  return(yint-shift+curve)
}
dereg <- function(x) {
  yint<- 11.037
  shift <- 0.7696*x
  curve <- 0.086*x*x
  return(yint-shift+curve)
}
```

Plotting

```r
summary(TRUCKING$DISTANCE)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.250   1.875   3.000   2.931   3.600   6.000
```

```r
x <- seq(0.25, 6, 0.25)
plot(x, reg(x),
     main = "Plot of PREDICT vs DISTANCE",
     ylab = "Predicted LN(PRICE)",
     xlab = "DISTANCE",
     ylim = range(9:12),
     type = "o",
     pch = "o",
     col = "darkmagenta")
points(x, dereg(x), col = "darkgreen", pch = "+")
lines(x, dereg(x), col = "darkgreen")
legend("topright", c("regulated", "deregulated"),
       fill = c("darkmagenta", "darkgreen"))
```



The graph clearly shows the impact of deregulation on the prices charged when the carrier leaves from Jacksonville with a cargo of 15,000 pounds. As expected from economic theory, the curve for the regulated prices lies above the curve for deregulated prices.

Follow-up Questions

1) In the Plot, give an expression (in terms of the estimated β's from Model 7) for the difference between the predicted regulated price and predicted deregulated price for any fixed value of mileage.

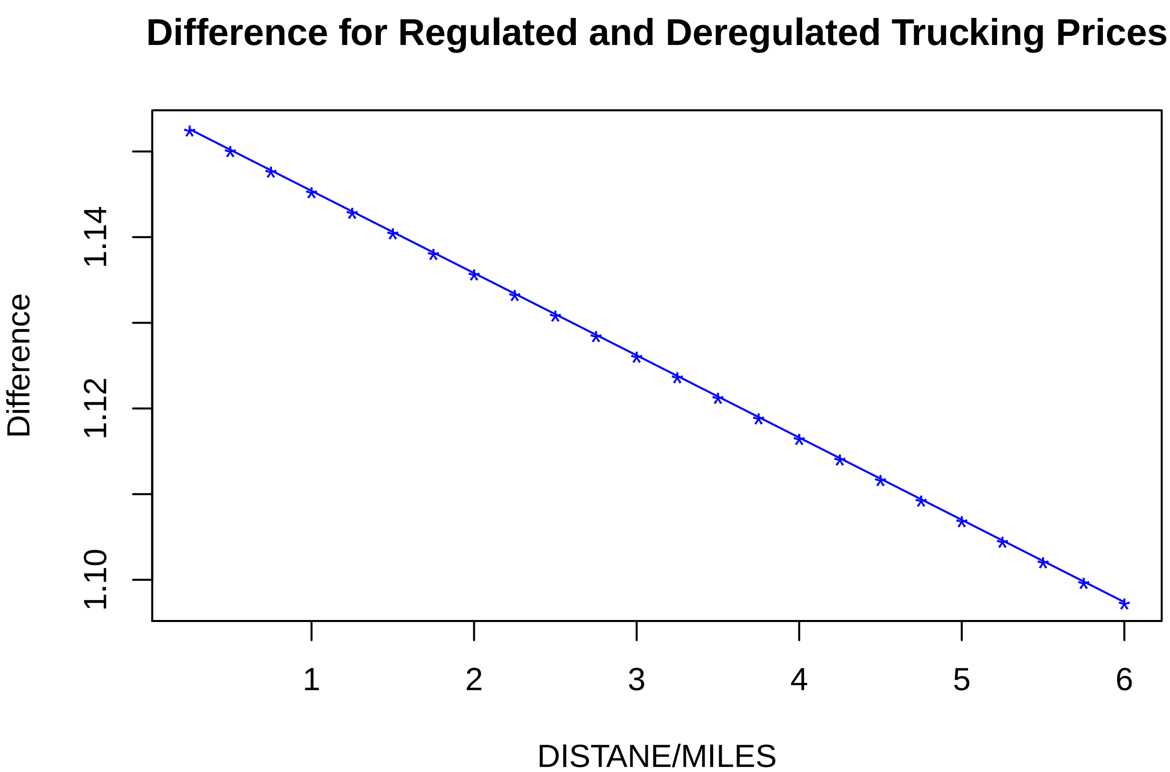
Regulated - Deregulated:

$$(12.192 - .760x_1 + .086x_1^2) - (11.037 - .7696x_1 + .086x_1^2)$$

$$= 1.155 - 0.0096x_1$$

```
diff <- function(x) {
  yint <- 1.155
  slope <- (0.0096*x)
  return(yint-slope)
}
```

```
plot(x, diff(x),
     main = "Difference for Regulated and Deregulated Trucking Prices",
     ylab = "Difference",
     xlab = "DISTANE/MILES",
     col = "blue",
     type = "o",
     pch = "*")
```



2) Demonstrate the impact of deregulation on price charged using the estimated β's from Model 7 in a fashion similar to the case study, but now hold origin fixed at Miami and weight fixed at 10,000 pounds.

Plugging in $x_2 = 10$ and $x_4 = 1$ we get:

$$\hat{y} = 12.192 - .598x_1 - .00598(10) - .01078x_1(10) + .086x_1^2 + .00014(10)^2$$

$$+.677(1) - .275x_1(1) - .026(10)(1) + .013x_1(10)(1) - .782x_3 + .0399x_1x_3$$

$$-.021(10)x_3 - .0033x_1(10)x_3$$

$$= 12.5632 - .8508x_1 + .086x_1^2 - .992x_3 + .0069x_1x_3$$

To see the impact of deregulation now we will plug in $x_3 = 1$ (deregulated) and $x_3 = 0$ (regulated), shown below:

$$\text{Regulated}(x_3 = 0) : \hat{y} = 12.5632 - .8508x_1 + .086x_1^2 - .992(0) + .0069x_1(0)$$

$$= 12.5632 - .8508x_1 + .086x_1^2$$

$$\text{Deregulation}(x_3 = 1) : \hat{y} = 12.5632 - .8508x_1 + .086x_1^2 - .992(1) + .0069x_1(1)$$

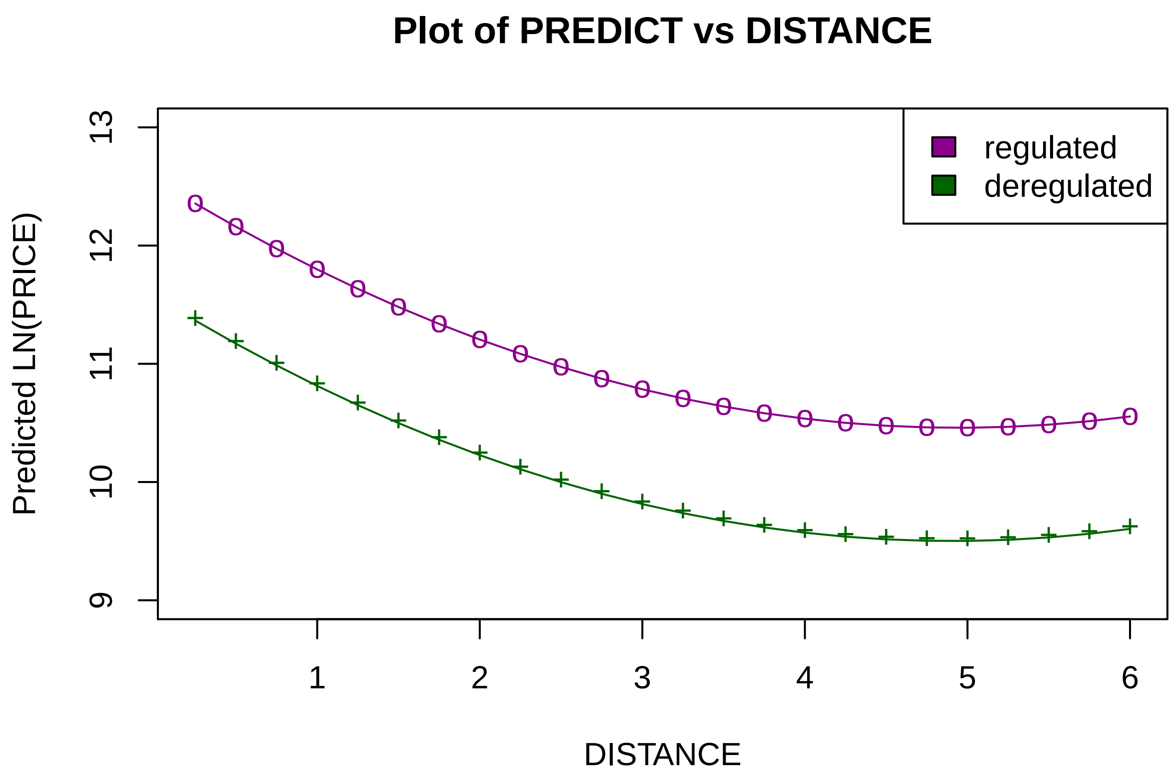
$$= 11.5712 - .8439x_1 + .086x_1^2$$

We can see that the y-intercept for the regulated prices is larger than the y-intercept for the deregulated prices. The equations have the same curvature but the shift parameter differs.

```r
regm <- function(x) {
  yint<- 12.5632
  shift <- 0.8508*x
  curve <- 0.086*x*x
  return(yint-shift+curve)
}
deregm <- function(x) {
  yint<- 11.5712
  shift <- 0.8439*x
  curve <- 0.086*x*x
  return(yint-shift+curve)
}
```

Plotting:

```r
plot(x, regm(x),
     main = "Plot of PREDICT vs DISTANCE",
     ylab = "Predicted LN(PRICE)",
     xlab = "DISTANCE",
     ylim = range(9:13),
     type = "o",
     pch = "o",
     col = "darkmagenta")
points(x, deregm(x), col = "darkgreen", pch = "+")
lines(x, deregm(x), col = "darkgreen")
legend("topright", c("regulated", "deregulated"),
       fill = c("darkmagenta", "darkgreen"))
```



The graph clearly shows the impact of deregulation on the prices charged when the carrier leaves from Miami with a cargo of 10,000 pounds. As expected from economic theory, the curve for the regulated prices lies above the curve for deregulated prices.

3) The data file TRUCKING4 contains data on trucking prices for four Florida carriers (A, B, C, and D). These carriers are identified by the variable CAR-RIER. (Note: Carrier B is the carrier analyzed in the case study.) Using Model 7 as a base model, add terms that allow for different response curves for the four carriers. Conduct the appropriate test to determine if the curves differ.

```
load("TRUCKING4.Rdata")
head(TRUCKING4)
```

```
##   PRICEPTM DISTANCE WEIGHT PCTLOAD   ORIGIN   MARKET    DEREG  CARRIER PRODUCT
## 1    31344     3.60   3.00    12.5 MIA      LARGE    YES      A            100
## 2   225676     0.25   0.75     3.1 MIA      LARGE    YES      A            100
## 3   172973     0.25   3.00    12.5 MIA      LARGE    YES      A            100
## 4    47167     2.60   0.25     1.0 MIA      LARGE    YES      A            100
## 5    30795     2.60  15.00    62.5 MIA      LARGE    YES      A            100
## 6    51126     1.50   3.00    12.5 MIA      SMALL    YES      A            100
##   LNPRICE
## 1 10.3528
## 2 12.3269
## 3 12.0609
## 4 10.7615
## 5 10.3351
## 6 10.8421
```

```
str(TRUCKING4)
```

```
## 'data.frame':    448 obs. of  10 variables:
##  $ PRICEPTM: num  31344 225676 172973 47167 30795 ...
```

```
##  $ DISTANCE: num  3.6 0.25 0.25 2.6 2.6 1.5 4.8 1.8 3.4 3.4 ...
##  $ WEIGHT  : num  3 0.75 3 0.25 15 3 15 15 7.5 24 ...
##  $ PCTLOAD : num  12.5 3.1 12.5 1 62.5 12.5 62.5 62.5 31.3 100 ...
##  $ ORIGIN  : Factor w/ 2 levels "JAX     ","MIA     ": 2 2 2 2 2 2 2 2 1 1 ...
##  $ MARKET  : Factor w/ 2 levels "LARGE   ","SMALL   ": 1 1 1 1 1 2 1 2 1 1 ...
##  $ DEREG   : Factor w/ 2 levels "NO      ","YES     ": 2 2 2 2 2 2 2 2 2 2 ...
##  $ CARRIER : Factor w/ 4 levels "A       ","B       ",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ PRODUCT : num  100 100 100 100 100 100 100 100 100 100 ...
##  $ LNPRICE : num  10.4 12.3 12.1 10.8 10.3 ...
```

Original Model 7 (Without Carrier Terms):

$$\hat{y} = 12.192 - .598x_1 - .00598x_2 - .01078x_1x_2 + .086x_1^2 + .00014x_2^2$$

$$+.677x_4 - .275x_1x_4 - .026x_2x_4 + .013x_1x_2x_4 - .782x_3$$

$$+.0399x_1x_3 - .021x_2x_3 - .0033x_1x_2x_3$$

Adding in x_5, x_6, x_7 for Carrier:

$$x_5 = \begin{cases} 1 & \text{if Carrier B} \\ 0 & \text{if not} \end{cases}$$

$$x_6 = \begin{cases} 1 & \text{if Carrier C} \\ 0 & \text{if not} \end{cases}$$

$$x_7 = \begin{cases} 1 & \text{if Carrier D} \\ 0 & \text{if not} \end{cases}$$

```
Y <- TRUCKING4$LNPRICE
X1 <- TRUCKING4$DISTANCE
X2 <- TRUCKING4$WEIGHT
X3 <- TRUCKING4$DEREG
X4 <- TRUCKING4$ORIGIN
X5 <- TRUCKING4$CARRIER
```

```
model8 <- lm(Y ~ X1 + X2 + X1*X2 + I(X1^2) + I(X2^2) + X3 + X4 + X5 + X1*X3 + X1*X4 + X1*X5 + X2*X3 + X2
summary(model8)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X1 * X2 + I(X1^2) + I(X2^2) + X3 +
##     X4 + X5 + X1 * X3 + X1 * X4 + X1 * X5 + X2 * X3 + X2 * X4 +
##     X2 * X5 + X1 * X2 * X3 + X1 * X2 * X4 + X1 * X2 * X5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.80936 -0.27121  0.01164  0.25445  1.10827
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       12.5184657  0.1884266  66.437  < 2e-16 ***
```

```
## X1               -0.6812502  0.0753746  -9.038  < 2e-16 ***
## X2               -0.0301213  0.0169919  -1.773 0.077003 .
## I(X1^2)           0.0851642  0.0078286  10.879  < 2e-16 ***
## I(X2^2)           0.0002683  0.0003202   0.838 0.402579
## X3YES            -0.4332804  0.1272635  -3.405 0.000726 ***
## X4MIA             0.6113791  0.1654084   3.696 0.000248 ***
## X5B              -0.4475470  0.1452336  -3.082 0.002194 **
## X5C              -0.2001413  0.1602276  -1.249 0.212319
## X5D              -0.2314838  0.1955857  -1.184 0.237261
## X1:X2            -0.0007936  0.0051284  -0.155 0.877096
## X1:X3YES          0.0437210  0.0416025   1.051 0.293895
## X1:X4MIA         -0.2306217  0.0556969  -4.141 4.18e-05 ***
## X1:X5B            0.0512131  0.0459311   1.115 0.265485
## X1:X5C            0.0606455  0.0497078   1.220 0.223131
## X1:X5D            0.0825943  0.0587831   1.405 0.160737
## X2:X3YES          0.0030368  0.0109045   0.278 0.780772
## X2:X4MIA         -0.0266516  0.0137604  -1.937 0.053433 .
## X2:X5B            0.0050117  0.0121126   0.414 0.679259
## X2:X5C            0.0133513  0.0157257   0.849 0.396358
## X2:X5D            0.0355065  0.0153800   2.309 0.021447 *
## X1:X2:X3YES      -0.0031538  0.0034096  -0.925 0.355514
## X1:X2:X4MIA       0.0090046  0.0046210   1.949 0.052001 .
## X1:X2:X5B        -0.0057077  0.0037031  -1.541 0.123986
## X1:X2:X5C        -0.0074617  0.0048222  -1.547 0.122524
## X1:X2:X5D        -0.0109979  0.0047146  -2.333 0.020131 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3638 on 422 degrees of freedom
## Multiple R-squared:  0.7426, Adjusted R-squared:  0.7273
## F-statistic: 48.69 on 25 and 422 DF,  p-value: < 2.2e-16
```

The End