Ex_4_Solution

Naga Soundari, Manasi, Mervin, Aditya

October 16, 2018

Load Data

```
# Read in the Data
priceData = read.csv("Ex4_Data_R.csv")

# Explore the data
str(priceData)
```

```
## 'data.frame':    66 obs. of  3 variables:
##  $ Day        : Factor w/ 66 levels "7/1/2017","7/10/2017",..: 1 12 23 26 27 28 29 30 31 2 ...
##  $ Visitors   : int  530 530 530 511 511 511 514 514 514 545 ...
##  $ Total.Spent: int  365 365 365 435 435 435 303 303 303 188 ...
```

```
summary(priceData)
```

```
##          Day        Visitors       Total.Spent
##  7/1/2017 : 1   Min.   : 511.0   Min.   : 188.0
##  7/10/2017: 1   1st Qu.: 615.0   1st Qu.: 430.0
##  7/11/2017: 1   Median : 918.5   Median : 786.5
##  7/12/2017: 1   Mean   : 862.1   Mean   : 714.1
##  7/13/2017: 1   3rd Qu.:1044.0   3rd Qu.: 884.0
##  7/14/2017: 1   Max.   :1197.0   Max.   :1400.0
##  (Other)  :60
```

PREPARE DATA

In our case we have to create new variables: log(Visitors), Visitors^2, and Lag(Day)

```
# Create log Advertising (we have done this for price so look it up!)
priceData$logSpent <- log(priceData$Total.Spent)

# Create Advertising^2
priceData$sqSpent <- (priceData$Total.Spent)^2

# Create LAG Sales
priceData$laggedVisitors <- c(NA, head(priceData$Visitors, -1))

# Display the data
summary(priceData)
```

```
##          Day        Visitors       Total.Spent        logSpent
##  7/1/2017 : 1   Min.   : 511.0   Min.   : 188.0   Min.   :5.236
##  7/10/2017: 1   1st Qu.: 615.0   1st Qu.: 430.0   1st Qu.:6.064
##  7/11/2017: 1   Median : 918.5   Median : 786.5   Median :6.667
##  7/12/2017: 1   Mean   : 862.1   Mean   : 714.1   Mean   :6.453
##  7/13/2017: 1   3rd Qu.:1044.0   3rd Qu.: 884.0   3rd Qu.:6.784
##  7/14/2017: 1   Max.   :1197.0   Max.   :1400.0   Max.   :7.244
```

```
##  (Other)  :60
##     sqSpent        laggedVisitors
##  Min.   :  35344   Min.   : 511.0
##  1st Qu.: 184900   1st Qu.: 615.0
##  Median : 618823   Median : 872.0
##  Mean   : 610178   Mean   : 860.5
##  3rd Qu.: 781456   3rd Qu.:1044.0
##  Max.   :1960000   Max.   :1197.0
##                    NA's   :1
```

Question 1: Short run response

REGRESSION MODELS

1) Define y (Dependent Variable). For this example, Visitors is used.
2) Define X (Independent Variables). For this example, Total.Spent is used.
3) Run the Models. For example, a linear model where No of visitors (Visitors) is a function of total money spent (Total.Spent)
   (a) Fit the following three models of advertising-click response (at this stage we will not add other information to the model; keep things simple and stick to advertising), report your results and comment, briefly.

```r
###Simple Linear
# Run the Regression (includes an INTERCEPT)
linearModel <- lm(Visitors ~ Total.Spent, data = priceData)
summary(linearModel)
```

```
##
## Call:
## lm(formula = Visitors ~ Total.Spent, data = priceData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -224.92 -158.68  -32.95  120.84  410.81
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 533.06454   52.36345  10.180  5.2e-15 ***
## Total.Spent   0.46076    0.06703   6.873  3.1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 172.4 on 64 degrees of freedom
## Multiple R-squared:  0.4247, Adjusted R-squared:  0.4157
## F-statistic: 47.24 on 1 and 64 DF,  p-value: 3.1e-09
```

```r
###Concave Logarithmic
# Run the Regression (includes an INTERCEPT)
concaveLogModel <- lm(Visitors ~ logSpent, data = priceData)
summary(concaveLogModel)
```

```
##
## Call:
```

```
## lm(formula = Visitors ~ logSpent, data = priceData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -237.60 -165.78   -8.76   97.31  396.87
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1075.92     255.38  -4.213 8.05e-05 ***
## logSpent      300.32      39.45   7.613 1.55e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 164.7 on 64 degrees of freedom
## Multiple R-squared:  0.4752, Adjusted R-squared:  0.467
## F-statistic: 57.95 on 1 and 64 DF,  p-value: 1.554e-10
```

```
###Concave Quadratic
# Run the Regression (includes an INTERCEPT)
concaveQuadModel <- lm(Visitors ~ Total.Spent + sqSpent, data = priceData)
summary(concaveQuadModel)
```

```
##
## Call:
## lm(formula = Visitors ~ Total.Spent + sqSpent, data = priceData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -226.79 -148.75   23.52   72.67  408.24
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.816e+02  9.675e+01   2.911  0.00498 **
## Total.Spent  1.293e+00  2.827e-01   4.574 2.29e-05 ***
## sqSpent     -5.621e-04  1.861e-04  -3.021  0.00364 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 162.4 on 63 degrees of freedom
## Multiple R-squared:  0.4975, Adjusted R-squared:  0.4815
## F-statistic: 31.18 on 2 and 63 DF,  p-value: 3.857e-10
```

(b) Compute the advertising elasticity implied by each model (use July's monthly clicks and advertising spent to scale the elasticities).

```
# Making a subset of July data
priceData.july <- priceData[(substr(priceData$Day,1,1) == '7'),]


meanVisitors <- mean(priceData.july$Visitors)
meanSpent <- mean(priceData.july$Total.Spent)

#linear model elasticity
elasticity_lm <- linearModel$coefficients[2] * (meanSpent/meanVisitors)
elasticity_lm
```

```
## Total.Spent
##   0.3981691
```

```
#Concave Logarithmic model elasticity
elasticity_conLog <- concaveLogModel$coefficients[2] * (1/meanVisitors)
elasticity_conLog
```

```
##  logSpent
## 0.4595617
```

```
#Concave Quadratic model elasticity
elasticity_conQuad <- (concaveQuadModel$coefficients[2] +
                        (2 * concaveQuadModel$coefficients[3]) * meanSpent) * (meanSpent/meanVisitors)
elasticity_conQuad
```

```
## Total.Spent
##   0.5689035
```

(c) Which model is best? Which, if any, would you reject?

Out of the three models, Linear regression model has the lowest R squared value. Also, going with the linear model could be a bit dicey for us because there are fewer data points, and the model doesn't fit accurately with fewer data points. Between the other two models, as both of them have similar Rsquared value, we would consider **Concave Logarithmic** model as it is simpler than the other. Also the F-statistic value is higher for Concave logarithmic model which means it explains more variance. In addition, the advertising models do not fit well with the concave quadratic model.

Question 2: Long run response

(a) Fit the same three models of advertising response but incorporate an exponentially decaying lag effect for advertising. Report your results and briefly comment.

```
###Linear with carry-over
# Run the Regression (includes an INTERCEPT)
linearModel_lag <- lm(Visitors ~ Total.Spent + laggedVisitors, data = priceData)
summary(linearModel_lag)
```

```
##
## Call:
## lm(formula = Visitors ~ Total.Spent + laggedVisitors, data = priceData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -97.371 -15.012  -6.624   7.324 159.681
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    38.50704   21.63573   1.780   0.0800 .
## Total.Spent     0.04992    0.02190   2.279   0.0261 *
## laggedVisitors  0.92129    0.03074  29.973   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.16 on 62 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9622, Adjusted R-squared:  0.9609
## F-statistic: 788.5 on 2 and 62 DF,  p-value: < 2.2e-16
```

```r
###Concave Logarithmic with carry-over
# Run the Regression (includes an INTERCEPT)
concaveLogModel_lag <- lm(Visitors ~ logSpent + laggedVisitors, data = priceData)
summary(concaveLogModel_lag)
```

```
##
## Call:
## lm(formula = Visitors ~ logSpent + laggedVisitors, data = priceData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -98.087 -17.810  -6.218   4.808 162.816
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -120.65169   76.61903  -1.575   0.1204
## logSpent        30.56469   14.19507   2.153   0.0352 *
## laggedVisitors   0.91847    0.03234  28.397   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.34 on 62 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9619, Adjusted R-squared:  0.9606
## F-statistic: 781.7 on 2 and 62 DF,  p-value: < 2.2e-16
```

```r
###Concave Quadratic with carry-over
# Run the Regression (includes an INTERCEPT)
concaveQuadModel_lag <- lm(Visitors ~ Total.Spent + sqSpent + laggedVisitors, data = priceData)
summary(concaveQuadModel_lag)
```

```
##
## Call:
## lm(formula = Visitors ~ Total.Spent + sqSpent + laggedVisitors,
##     data = priceData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -97.582 -15.634  -6.140   6.599 159.999
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.541e+01  2.831e+01   1.251    0.216
## Total.Spent     6.474e-02  8.911e-02   0.727    0.470
## sqSpent        -9.391e-06  5.470e-05  -0.172    0.864
## laggedVisitors  9.192e-01  3.320e-02  27.686   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.51 on 61 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9622, Adjusted R-squared:  0.9603
## F-statistic: 517.4 on 3 and 61 DF,  p-value: < 2.2e-16
```

(b) Compute the long-run advertising elasticity implied by each model (again use July's monthly figures for scaling).

```
#linear model elasticity
elasticity_lm_lag <- (linearModel_lag$coefficients[2]/(1 -
                                                          (linearModel_lag$coefficients[3]))) *
  (meanSpent/meanVisitors)
elasticity_lm_lag
```

```
## Total.Spent
##   0.5480682
```

```
#Concave Logarithmic model elasticity
elasticity_conLog_lag <- (concaveLogModel_lag$coefficients[2]/(1 -
(concaveLogModel_lag$coefficients[3]))) *
 (1/meanVisitors)
elasticity_conLog_lag
```

```
##  logSpent
## 0.5736757
```

```
#Concave Quadratic model elasticity
elasticity_conQuad_lag <- ((concaveQuadModel_lag$coefficients[2] +
                               (2 * concaveQuadModel_lag$coefficients[3]) * meanSpent)/(1 -
                                   (concaveQuadModel_lag$coefficients[4]))) *
  (meanSpent/meanVisitors)
elasticity_conQuad_lag
```

```
## Total.Spent
##   0.5792549
```

(c) Which model is best? Which, if any, would you reject?

Running the three models with carryover effects which is the result of immediate plus lagged values, we get the same results in all the three models. The way to justify the reliability of the model is to look at the R squared and the p value. Looking at the three, we see that the adjusted R squared is ~0.9 and for the p-value, all the co-efficients in Linear model have significant values compared to the other models which talk more about the reliability of the model. Also stating the fact, that Linear model is the simplest, it would be a wise choice to go with the **Linear model**.

If there is any model we need to reject, Concave Quadratic would be the one. Reason being, the significance of the co-efficients is least observed in Concave Quad.

Question 3: Saturation.

Drawing on the models that you fitted above, compute the saturation level for advertising spending. Report results in terms of daily advertising spending.

```
saturationWithoutLag <- -(concaveQuadModel$coefficients[2]/(2 * concaveQuadModel$coefficients[3]))
saturationWithoutLag
```

```
## Total.Spent
##    1150.322
```

```
saturationWithLag <- -(concaveQuadModel_lag$coefficients[2]/(2 * concaveQuadModel_lag$coefficients[3]))
saturationWithLag
```

```
## Total.Spent
##    3447.066
```

The saturation level indiciates the cap amount or threshold. We applied the formula (-(b1/2b2)) and looking at the value without the carryover, we can conclude that the model does good as long as it's advertising investment is **1150 or under per day**. The same goes for value with carryover. As carryover also carries the lag with it, the advertising investment can't be more than **3447 per day**. If it is, we would say that our model/strategy is not doing good.