Cloverleaf Regression Analysis

Danny Moncada (monca016), MSBA

```
## Load data file
data <- read.csv("PaidSearch.csv", header = T)

# Do variables (adrank1) play a role
data$adrank1 <- ifelse(data$adrank == 1, 1, 0)

attach(data)

### CONVERSION RATE ###

# Do variables ( as is) play a role
ConRateReg1 <- lm(conversionrate ~ clicks + clickthroughrate + adrank + numberofwords + retailer + brand
summary(ConRateReg1)
```

```
##
## Call:
## lm(formula = conversionrate ~ clicks + clickthroughrate + adrank +
##     numberofwords + retailer + brandname + landQuality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21087 -0.05044 -0.02140  0.01930  1.01359
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -6.803e-02  1.726e-02  -3.941 8.68e-05 ***
## clicks           -5.410e-06  2.930e-06  -1.847   0.0651 .
## clickthroughrate -1.912e-01  3.119e-02  -6.131 1.25e-09 ***
## adrank            2.907e-03  4.968e-04   5.852 6.56e-09 ***
## numberofwords     8.685e-03  4.872e-03   1.783   0.0750 .
## retailer          1.031e-01  1.420e-02   7.261 7.68e-13 ***
## brandname         5.850e-02  1.370e-02   4.269 2.15e-05 ***
## landQuality       4.633e-03  1.765e-03   2.625   0.0088 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1532 on 1002 degrees of freedom
## Multiple R-squared:  0.1082, Adjusted R-squared:  0.1019
## F-statistic: 17.36 on 7 and 1002 DF,  p-value: < 2.2e-16
```

```
ConRateReg2 <- lm(conversionrate ~ clicks + clickthroughrate + adrank1 + numberofwords + retailer + bra
summary(ConRateReg2)
```

```
##
## Call:
## lm(formula = conversionrate ~ clicks + clickthroughrate + adrank1 +
##     numberofwords + retailer + brandname + landQuality)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18458 -0.05948 -0.01277  0.00791  1.02189
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.267e-02  1.443e-02  -0.878 0.380133
## clicks           -4.614e-06  2.952e-06  -1.563 0.118285
## clickthroughrate -1.194e-01  3.769e-02  -3.168 0.001579 **
## adrank1          -6.880e-02  1.469e-02  -4.683 3.22e-06 ***
## numberofwords     8.611e-03  4.916e-03   1.752 0.080133 .
## retailer          9.610e-02  1.450e-02   6.626 5.63e-11 ***
## brandname         5.101e-02  1.365e-02   3.737 0.000197 ***
## landQuality       3.400e-03  1.761e-03   1.931 0.053792 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1541 on 1002 degrees of freedom
## Multiple R-squared:  0.09743,    Adjusted R-squared:  0.09113
## F-statistic: 15.45 on 7 and 1002 DF,  p-value: < 2.2e-16
```

```
### CLICK THROUGH RATE ###

# Do variables ( as is) play a role
ClickThroughReg1 <- lm(clickthroughrate ~ adrank + numberofwords + retailer + brandname + landQuality)
summary(ClickThroughReg1)
```

```
##
## Call:
## lm(formula = clickthroughrate ~ adrank + numberofwords + retailer +
##     brandname + landQuality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.53349 -0.07299 -0.00998  0.09263  1.00671
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.075070   0.017365   4.323 1.69e-05 ***
## adrank        -0.002533   0.000500  -5.067 4.81e-07 ***
## numberofwords -0.041242   0.004737  -8.706  < 2e-16 ***
## retailer       0.287253   0.011183  25.686  < 2e-16 ***
## brandname      0.087186   0.013692   6.368 2.92e-10 ***
## landQuality    0.008937   0.001777   5.030 5.80e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1561 on 1004 degrees of freedom
## Multiple R-squared:  0.5385, Adjusted R-squared:  0.5362
## F-statistic: 234.3 on 5 and 1004 DF,  p-value: < 2.2e-16
```

```
# Do variables (adrank1) play a role
ClickThroughReg2 <- lm(clickthroughrate ~ adrank1 + numberofwords + retailer + brandname + landQuality)
summary(ClickThroughReg2)
```

```
##
## Call:
## lm(formula = clickthroughrate ~ adrank1 + numberofwords + retailer +
##     brandname + landQuality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39938 -0.09888  0.01300  0.07166  0.90391
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.021841   0.012031   1.815   0.0698 .
## adrank1        0.224432   0.010072  22.284  < 2e-16 ***
## numberofwords -0.037144   0.003915  -9.488  < 2e-16 ***
## retailer       0.233994   0.009585  24.411  < 2e-16 ***
## brandname      0.006174   0.011452   0.539   0.5899
## landQuality    0.006597   0.001463   4.509 7.28e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1293 on 1004 degrees of freedom
## Multiple R-squared:  0.6834, Adjusted R-squared:  0.6818
## F-statistic: 433.4 on 5 and 1004 DF,  p-value: < 2.2e-16
```

```
### REVENUE ###

# Do variables ( as is) play a role
RevReg1 <- lm(revenue ~ clickthroughrate + adrank + numberofwords + retailer + brandname + landQuality)
summary(RevReg1)
```

```
##
## Call:
## lm(formula = revenue ~ clickthroughrate + adrank + numberofwords +
##     retailer + brandname + landQuality)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2097.6  -606.9  -258.3    19.5  9075.4
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -550.37570  201.85447  -2.727 0.006511 **
## clickthroughrate 1299.43995  363.48462   3.575 0.000367 ***
## adrank              0.01538    5.83187   0.003 0.997896
## numberofwords     260.90849   56.58253   4.611 4.52e-06 ***
## retailer          569.41413  165.80555   3.434 0.000619 ***
## brandname           7.72432  160.85023   0.048 0.961708
## landQuality         8.14088   20.71815   0.393 0.694451
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1798 on 1003 degrees of freedom
## Multiple R-squared:  0.1001, Adjusted R-squared:  0.09472
## F-statistic: 18.59 on 6 and 1003 DF,  p-value: < 2.2e-16
```

```
# Do variables (adrank1) play a role
RevReg2 <- lm(revenue ~ clickthroughrate + adrank1 + numberofwords + retailer + brandname + landQuality)
summary(RevReg2)
```

```
##
## Call:
## lm(formula = revenue ~ clickthroughrate + adrank1 + numberofwords +
##     retailer + brandname + landQuality)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1944.5  -685.2  -246.7   111.4  9011.0
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -544.32     167.23  -3.255  0.00117 **
## clickthroughrate   809.94     437.98   1.849  0.06471 .
## adrank1            331.88     170.87   1.942  0.05238 .
## numberofwords      245.01      56.71   4.320 1.71e-05 ***
## retailer           629.09     167.92   3.746  0.00019 ***
## brandname          -97.70     158.95  -0.615  0.53891
## landQuality          7.65      20.51   0.373  0.70924
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1795 on 1003 degrees of freedom
## Multiple R-squared:  0.1035, Adjusted R-squared:  0.09811
## F-statistic: 19.29 on 6 and 1003 DF,  p-value: < 2.2e-16
```

```
#with converstion rate included
# Do variables ( as is) play a role
RevReg3 <- lm(revenue ~ clickthroughrate + adrank + numberofwords + retailer + brandname + landQuality +
summary(RevReg3)
```

```
##
## Call:
## lm(formula = revenue ~ clickthroughrate + adrank + numberofwords +
##     retailer + brandname + landQuality + conversionrate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2085.7  -595.4  -237.8    46.4  9049.9
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -586.723    203.165  -2.888 0.003961 **
## clickthroughrate 1188.954    370.584   3.208 0.001377 **
## adrank              1.634      5.926   0.276 0.782797
## numberofwords     265.016     56.612   4.681 3.24e-06 ***
```

```
## retailer          625.403    169.817   3.683 0.000243 ***
## brandname          40.385    162.204   0.249 0.803429
## landQuality        10.695     20.774   0.515 0.606802
## conversionrate   -557.412    370.008  -1.506 0.132258
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1797 on 1002 degrees of freedom
## Multiple R-squared:  0.1021, Adjusted R-squared:  0.09586
## F-statistic: 16.28 on 7 and 1002 DF,  p-value: < 2.2e-16
```

```
# Do variables (adrank1) play a role
RevReg4 <- lm(revenue ~ clickthroughrate + adrank1 + numberofwords + retailer + brandname + landQuality
summary(RevReg4)
```

```
##
## Call:
## lm(formula = revenue ~ clickthroughrate + adrank1 + numberofwords +
##     retailer + brandname + landQuality + conversionrate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1952.9  -669.2  -232.1    85.4  8994.1
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -548.90     167.24  -3.282  0.00107 **
## clickthroughrate   755.13     440.22   1.715  0.08659 .
## adrank1            300.82     172.76   1.741  0.08195 .
## numberofwords      248.36      56.77   4.375 1.34e-05 ***
## retailer           670.62     171.37   3.913 9.72e-05 ***
## brandname          -74.85     160.04  -0.468  0.64008
## landQuality          9.14      20.54   0.445  0.65646
## conversionrate    -443.62     367.44  -1.207  0.22758
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1795 on 1002 degrees of freedom
## Multiple R-squared:  0.1048, Adjusted R-squared:  0.09852
## F-statistic: 16.75 on 7 and 1002 DF,  p-value: < 2.2e-16
```

```
### Separate data into Retailer and Non ##
dataRetailer <- data[retailer == 1,]
dataNonRetailer <- data[retailer == 0,]


### CONVERSION RATE ###

#Retailer
# Do variables ( as is) play a role
RetConRateReg1 <- lm(conversionrate ~ clicks + clickthroughrate + adrank + numberofwords + brandname + l
summary(RetConRateReg1)
```

```
##
```

```
## Call:
## lm(formula = conversionrate ~ clicks + clickthroughrate + adrank +
##     numberofwords + brandname + landQuality, data = dataRetailer)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70795 -0.02046 -0.00033  0.00633  0.65356
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -3.697e-01  4.188e-02  -8.828   <2e-16 ***
## clicks            6.993e-07  2.944e-06   0.238    0.812
## clickthroughrate -4.776e-02  4.258e-02  -1.122    0.263
## adrank            3.551e-02  1.793e-03  19.803   <2e-16 ***
## numberofwords    -5.160e-03  7.142e-03  -0.722    0.471
## brandname         3.899e-01  3.798e-02  10.267   <2e-16 ***
## landQuality      -1.917e-03  3.533e-03  -0.543    0.588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1457 on 361 degrees of freedom
## Multiple R-squared:  0.6095, Adjusted R-squared:  0.603
## F-statistic: 93.92 on 6 and 361 DF,  p-value: < 2.2e-16
```

```
# Do variables (adrank1) play a role
dataRetailer$adrank1 <- ifelse(dataRetailer$adrank == 1, 1, 0)
RetConRateReg2 <- lm(conversionrate ~ clicks + clickthroughrate + adrank1 + numberofwords + brandname +
summary(ConRateReg2)
```

```
##
## Call:
## lm(formula = conversionrate ~ clicks + clickthroughrate + adrank1 +
##     numberofwords + retailer + brandname + landQuality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18458 -0.05948 -0.01277  0.00791  1.02189
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.267e-02  1.443e-02  -0.878 0.380133
## clicks           -4.614e-06  2.952e-06  -1.563 0.118285
## clickthroughrate -1.194e-01  3.769e-02  -3.168 0.001579 **
## adrank1          -6.880e-02  1.469e-02  -4.683 3.22e-06 ***
## numberofwords     8.611e-03  4.916e-03   1.752 0.080133 .
## retailer          9.610e-02  1.450e-02   6.626 5.63e-11 ***
## brandname         5.101e-02  1.365e-02   3.737 0.000197 ***
## landQuality       3.400e-03  1.761e-03   1.931 0.053792 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1541 on 1002 degrees of freedom
## Multiple R-squared:  0.09743,    Adjusted R-squared:  0.09113
## F-statistic: 15.45 on 7 and 1002 DF,  p-value: < 2.2e-16
```

```r
#Non-retailer
# Do variables ( as is) play a role
NonConRateReg1 <- lm(conversionrate ~ clicks + clickthroughrate + adrank + numberofwords + brandname + l
summary(NonConRateReg1)
```

```
##
## Call:
## lm(formula = conversionrate ~ clicks + clickthroughrate + adrank +
##     numberofwords + brandname + landQuality, data = dataNonRetailer)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09102 -0.02312 -0.00599  0.00523  0.95831
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -5.735e-02  1.582e-02  -3.626 0.000311 ***
## clicks           -5.469e-07  1.006e-05  -0.054 0.956678
## clickthroughrate  2.763e-02  3.330e-02   0.830 0.406897
## adrank            4.786e-04  3.412e-04   1.403 0.161167
## numberofwords     2.193e-02  5.281e-03   4.153 3.73e-05 ***
## brandname        -1.586e-02  9.492e-03  -1.671 0.095242 .
## landQuality       5.944e-03  1.458e-03   4.077 5.15e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09694 on 635 degrees of freedom
## Multiple R-squared:  0.05111,    Adjusted R-squared:  0.04215
## F-statistic: 5.701 on 6 and 635 DF,  p-value: 8.706e-06
```

```r
# Do variables (adrank1) play a role
dataNonRetailer$adrank1 <- ifelse(dataNonRetailer$adrank == 1, 1, 0)
nonConRateReg2 <- lm(conversionrate ~ clicks + clickthroughrate + adrank1 + numberofwords + brandname +
summary(ConRateReg2)
```

```
##
## Call:
## lm(formula = conversionrate ~ clicks + clickthroughrate + adrank1 +
##     numberofwords + retailer + brandname + landQuality)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18458 -0.05948 -0.01277  0.00791  1.02189
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.267e-02  1.443e-02  -0.878 0.380133
## clicks           -4.614e-06  2.952e-06  -1.563 0.118285
## clickthroughrate -1.194e-01  3.769e-02  -3.168 0.001579 **
## adrank1          -6.880e-02  1.469e-02  -4.683 3.22e-06 ***
## numberofwords     8.611e-03  4.916e-03   1.752 0.080133 .
## retailer          9.610e-02  1.450e-02   6.626 5.63e-11 ***
```

```
## brandname         5.101e-02  1.365e-02   3.737 0.000197 ***
## landQuality       3.400e-03  1.761e-03   1.931 0.053792 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1541 on 1002 degrees of freedom
## Multiple R-squared:  0.09743,    Adjusted R-squared:  0.09113
## F-statistic: 15.45 on 7 and 1002 DF,  p-value: < 2.2e-16
```

```
### CLICK THROUGH RATE ###

#Retailer
# Do variables ( as is) play a role
RetClickThroughReg1 <- lm(clickthroughrate ~ adrank + numberofwords + brandname + landQuality, data = da
summary(RetClickThroughReg1)
```

```
##
## Call:
## lm(formula = clickthroughrate ~ adrank + numberofwords + brandname +
##     landQuality, data = dataRetailer)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.51741 -0.09204  0.07531  0.12162  0.56151
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.317837   0.048863   6.505 2.58e-10 ***
## adrank        -0.019568   0.001936 -10.108  < 2e-16 ***
## numberofwords -0.006380   0.008442  -0.756   0.4503
## brandname      0.213193   0.045350   4.701 3.68e-06 ***
## landQuality   -0.009106   0.004292  -2.122   0.0345 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1797 on 363 degrees of freedom
## Multiple R-squared:  0.3562, Adjusted R-squared:  0.3491
## F-statistic: 50.21 on 4 and 363 DF,  p-value: < 2.2e-16
```

```
# Do variables (adrank1) play a role
RetClickThroughReg2 <- lm(clickthroughrate ~ adrank1 + numberofwords + brandname + landQuality, data = d
summary(RetClickThroughReg2)
```

```
##
## Call:
## lm(formula = clickthroughrate ~ adrank1 + numberofwords + brandname +
##     landQuality, data = dataRetailer)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44209 -0.10331  0.01865  0.08223  0.41648
##
## Coefficients:
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.057862   0.031852   1.817   0.0701 .
## adrank1        0.434638   0.018186  23.900  < 2e-16 ***
## numberofwords -0.037258   0.005852  -6.367 5.82e-10 ***
## brandname      0.036008   0.033204   1.084   0.2789
## landQuality    0.006865   0.003023   2.271   0.0237 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1268 on 363 degrees of freedom
## Multiple R-squared:  0.6794, Adjusted R-squared:  0.6759
## F-statistic: 192.3 on 4 and 363 DF,  p-value: < 2.2e-16
```

```
#NonRetailer
# Do variables ( as is) play a role
NonClickThroughReg1 <- lm(clickthroughrate ~ adrank + numberofwords + brandname + landQuality, data = da
summary(NonClickThroughReg1)
```

```
##
## Call:
## lm(formula = clickthroughrate ~ adrank + numberofwords + brandname +
##     landQuality, data = dataNonRetailer)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18885 -0.05982 -0.03055  0.04228  1.00153
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.0188935  0.0187889   1.006    0.315
## adrank        -0.0007696  0.0004069  -1.891    0.059 .
## numberofwords -0.0301798  0.0059847  -5.043 5.99e-07 ***
## brandname      0.0665641  0.0110371   6.031 2.76e-09 ***
## landQuality    0.0145321  0.0016259   8.938  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1161 on 637 degrees of freedom
## Multiple R-squared:  0.2731, Adjusted R-squared:  0.2685
## F-statistic: 59.82 on 4 and 637 DF,  p-value: < 2.2e-16
```

```
# Do variables (adrank1) play a role
nonClickThroughReg2 <- lm(clickthroughrate ~ adrank1 + numberofwords + brandname + landQuality, data = d
summary(nonClickThroughReg2)
```

```
##
## Call:
## lm(formula = clickthroughrate ~ adrank1 + numberofwords + brandname +
##     landQuality, data = dataNonRetailer)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21806 -0.04527 -0.01181  0.03463  0.93997
```

```
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.006235   0.013717   0.455   0.6496
## adrank1        0.123232   0.009866  12.490  < 2e-16 ***
## numberofwords -0.028320   0.005185  -5.462 6.75e-08 ***
## brandname      0.020816   0.010231   2.035   0.0423 *
## landQuality    0.011536   0.001446   7.979 6.87e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1044 on 637 degrees of freedom
## Multiple R-squared:  0.4128, Adjusted R-squared:  0.4091
## F-statistic:   112 on 4 and 637 DF,  p-value: < 2.2e-16
```

```
## Release yourself from the data
detach(data)
```