ebay Marketing Assesment

Roopa, Uros

October 29, 2017

R Markdown

Load the dataset in R and create a new column called 'rDate' convert the 'date' column into the 'date' datatype.

```
#install.packages("anytime")
library("anytime")
```

```
## Warning: package 'anytime' was built under R version 3.3.3
```

```
library("zoo")
```

```
## Warning: package 'zoo' was built under R version 3.3.3
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
dataBay<-read.csv("C:\\Users\\Uros Randelovic\\Documents\\R workspace\\BUS 111\\Ebay\\data.csv")

dataBay$rDate <- as.Date(dataBay$date, format="%m/%d/%Y")

head(dataBay)
```

```
##       date dma isTreatmentPeriod isTreatmentGroup    revenue      rDate
## 1 4/1/2012 500                 0                0   76718.74 2012-04-01
## 2 4/1/2012 501                 0                1 2096176.54 2012-04-01
## 3 4/1/2012 502                 0                1   34993.85 2012-04-01
## 4 4/1/2012 503                 0                1   34198.75 2012-04-01
## 5 4/1/2012 504                 0                1  641014.21 2012-04-01
## 6 4/1/2012 505                 0                0  327989.31 2012-04-01
```

b) Determine the date that started the treatment period. That is, write code to determine the earliest date in the treatment period.

```
treatedPeriod <- dataBay[which(dataBay$isTreatmentPeriod==1),]
untreatedPeriod <-dataBay[which(dataBay$isTreatmentPeriod==0),]

#finding date of first treatment period
min(treatedPeriod$rDate)
```

```
## [1] "2012-05-22"
```

How were the treatment and control groups treated differently during this period? 5% - Treatment group was not shown ads in the time indicated and the control group was shown ads to determine the if the revenues will decrease.

c) The data contains a control group, which are shown search ads throughout the data, and a treatment group, which are only shown search ads before the treatment period.

i. Run a regression that compares log(revenue) of the treatment group in the pre-treatment period and in the treatment period. 5%

```
#create dataframe that contains only rows containing treated group - the group that ads were not shown
TreatmentGroupOnly = dataBay[which(dataBay$isTreatmentGroup==1),]
#log revenue of treatment group regressed on pre and post
summary(lm(log(revenue) ~ factor(isTreatmentPeriod), data=TreatmentGroupOnly))
```

```
##
## Call:
## lm(formula = log(revenue) ~ factor(isTreatmentPeriod), data = TreatmentGroupOnly)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0038 -0.7490 -0.0274  0.6929  3.8268
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 10.94865    0.01472 743.988   <2e-16 ***
## factor(isTreatmentPeriod)1 -0.03940    0.01987  -1.983   0.0474 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.252 on 16044 degrees of freedom
## Multiple R-squared:  0.0002451,  Adjusted R-squared:  0.0001828
## F-statistic: 3.933 on 1 and 16044 DF,  p-value: 0.04737
```

ii. What do the resulting coefficient estimates say about the effectiveness of advertising? Be as specific as you can. 10%

B1 is -0.04. This means that a unit change in the treatment period i.e whether the treatment group has been shown ads or not been shown ads will lead to a 0.041% change in their revenue. We are 99% confident about this.

d) Now we will use the control group for a true experimental approach. First, we will check if the randomization was done properly.

i. Run a regression that compares log(revenue) of the treatment group and the control group in the pre-treatment period. 10%

```
#extract only pretreatment period
preTreatmentOnly= dataBay[which(dataBay$isTreatmentPeriod==0),]

summary(lm(log(revenue)~ isTreatmentGroup,data = preTreatmentOnly))
```

```
##
## Call:
## lm(formula = log(revenue) ~ isTreatmentGroup, data = preTreatmentOnly)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.9962 -0.7502 -0.0285  0.7331  3.8229
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      10.96273    0.02037 538.128   <2e-16 ***
## isTreatmentGroup -0.01408    0.02477  -0.568     0.57
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.2 on 10708 degrees of freedom
## Multiple R-squared:  3.017e-05,  Adjusted R-squared:  -6.322e-05
## F-statistic: 0.323 on 1 and 10708 DF,  p-value: 0.5698
```

ii. What is the purpose of this randomization check? What do the results of this regression show? 5%

The randomization check seeks to make sure the treatment group or the control group are not generating a significant difference in the revenue before the treatment is even applied. That way, after the treatment, we can be sure that any difference in revenue can be attributed to the treatment and not some other factor present before. The results show that a unit change may cause a 0.01% change in revenue but that is not significant, which proves that we can attribute the treatment as the cause for our future results.

e) Now, using the post treatment data, determine the effectiveness of eBay ads.

i. Run a regression with log(revenue) as the dependent variable, and whether the DMA is in the treatment group as the independent variable. 10%

```
#add column to indetify all dma's that are in the treatment group and in treatment time (1 if the dma w
dataBay$adServed <- ifelse((dataBay$isTreatmentGroup==1 & dataBay$isTreatmentPeriod==1),1,0)

summary(lm(log(dataBay$revenue) ~ dataBay$adServed))
```

```
##
## Call:
## lm(formula = log(dataBay$revenue) ~ dataBay$adServed)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0038 -0.7506 -0.0285  0.7361  3.8287
##
## Coefficients:
##                   Estimate Std. Error  t value Pr(>|t|)
## (Intercept)      10.942905   0.009859 1109.934   <2e-16 ***
## dataBay$adServed -0.033659   0.016186   -2.079   0.0376 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.205 on 23728 degrees of freedom
## Multiple R-squared:  0.0001822,  Adjusted R-squared:  0.0001401
## F-statistic: 4.324 on 1 and 23728 DF,  p-value: 0.03758
```

ii. What do the resulting coefficient estimates say about the effectiveness of advertising? Be as specific as you can. 10%

When the advertising was stopped in the treatment period, we observe a decrease of 3% in revenues thus ads have some although a minor effect.

iii. What was the purpose of the control group here? What factor was unaccounted for in part c, but was accounted for in part e? 10%

Here we are purely looking at the post treatment group in the posttreatment period wheareas in the part c we were looking at the treatment group and comapring the revenues in the pre and post treatment periods. (there is no control group included in this regression per instructions (post treatment period, if dma is in the treated group) but the control group helped us previously determine if the assigment was truly random and show that our data is not skewed thus allowing for the correct inferencies)

iv. Does this R-squared of this regression affect the interpretation or confidence in the estimate of the effectiveness of advertising? 10%

Just because effect size is small doesn't mean it's bad, unworthy of being interpreted, or useless. It's just small. Even small effect sizes can have scientific sifnificance especially here that we are using log values in the regression.

f) Check whether the effectiveness of eBay advertising changes over time. Create a new variable called month, which contains a factor variable of the month that the date falls in. In a new regression on the post treatment data, interact whether the DMA is in the treatment group with this month variable. Paste a summary of this regression into your pdf. 10%

Hint: Creating the month variable can be done in a least two ways. One method would use the paste and the substr function. Another would use the lubridate package. Any approach that works is acceptable.

```
postTreatment <- dataBay[which(dataBay$isTreatmentPeriod==1),]

library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.3.3
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
dataBay$dateM <- month(dataBay$rDate)

summary(lm(dataBay$isTreatmentGroup~dataBay$dateM))
```

```
##
## Call:
## lm(formula = dataBay$isTreatmentGroup ~ dataBay$dateM)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6762 -0.6762  0.3238  0.3238  0.3238
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.762e-01  1.551e-02   43.59   <2e-16 ***
## dataBay$dateM -4.025e-15  2.823e-03    0.00        1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4679 on 23728 degrees of freedom
## Multiple R-squared:  4.306e-26,  Adjusted R-squared:  -4.214e-05
## F-statistic: 1.022e-21 on 1 and 23728 DF,  p-value: 1
```

g) Throughout the analysis regression were run on log(revenue) rather than revenue. Was this the right choice? Or would simply using revenue be more appropriate? Justify your answer. 10%

Substantively, sometimes the meaning of a change in a variable is more multiplicative than additive. For example, income. If you make 20,000 a year, a 5,000 raise is huge. If you make $200,000 a year, it is small. Taking logs reflects this thus looking at the incremental changes after the advertising has been stopped helps us understand the extent to which it had the effect.