MKT 436(R) - Homework 1 Solution

Haviv, Avery

Question 1A:

```
# Load Dataset
setwd('D:/Dropbox/Teaching Lectures/Assignments/Homework 1')
homeworkDB = read.csv('Homework 1 Data - 436.csv')
```

Question 1B:

```
#####################################################
# Load Dataset
homeworkDB[1:10,]
```

```
##        date dma isTreatmentPeriod isTreatmentGroup    revenue
## 1  4/1/2012 500                 0                0   76718.74
## 2  4/1/2012 501                 0                1 2096176.54
## 3  4/1/2012 502                 0                1   34993.85
## 4  4/1/2012 503                 0                1   34198.75
## 5  4/1/2012 504                 0                1  641014.21
## 6  4/1/2012 505                 0                0  327989.31
## 7  4/1/2012 506                 0                1  542373.98
## 8  4/1/2012 507                 0                1   49080.20
## 9  4/1/2012 508                 0                1  230705.29
## 10 4/1/2012 509                 0                1   51699.45
```

Question 1C:

```
#####################################################
#One approach
treatPeriod = subset(homeworkDB,homeworkDB$isTreatmentPeriod==1)
sort(treatPeriod$date)[1]
```

```
## [1] 5/22/2012
## 113 Levels: 4/1/2012 4/10/2012 4/11/2012 4/12/2012 4/13/2012 ... 7/9/2012
```

```
#Another approach
#Convert to proper dates
homeworkDB$rDate= as.Date(paste(homeworkDB$date),'%m/%d/%Y')
treatPeriod = subset(homeworkDB,homeworkDB$isTreatmentPeriod==1)

#What was the earliest treatment period?
min(treatPeriod$rDate)
```

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

Question 1D:

```
#No control group
treatOnly = subset(homeworkDB,homeworkDB$isTreatmentGroup==1)
summary(lm(log(revenue)~isTreatmentPeriod,data=treatOnly))
```

```
##
## Call:
## lm(formula = log(revenue) ~ isTreatmentPeriod, data = treatOnly)
##
## 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 ***
## isTreatmentPeriod -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
```

Question 1E:

```
#Randomization Check
preTreatment = subset(homeworkDB,homeworkDB$isTreatmentPeriod==0)
summary(lm(log(revenue)~isTreatmentGroup,data=preTreatment))
```

```
##
## Call:
## lm(formula = log(revenue) ~ isTreatmentGroup, data = preTreatment)
##
## 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
```

Question 1F:

```
#Part 1F: Run LM to figure out ad effects
postTreatment = subset(homeworkDB,homeworkDB$isTreatmentPeriod==1)
summary(lm(log(revenue)~isTreatmentGroup,data=postTreatment))
```

```
##
## Call:
## lm(formula = log(revenue) ~ isTreatmentGroup, data = postTreatment)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -5.0038 -0.7546 -0.0288  0.7419  3.8268
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      10.916740   0.018610 586.595   <2e-16 ***
## isTreatmentGroup -0.007494   0.022632  -0.331    0.741
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.208 on 13018 degrees of freedom
## Multiple R-squared:  8.422e-06,  Adjusted R-squared:  -6.839e-05
## F-statistic: 0.1096 on 1 and 13018 DF,  p-value: 0.7406
```

Question 2A: The coefficient estimates imply that the presence of advertising increases revenue by roughly 4%. While the data rejects the notion that advertising has no effect, ads might plausibly increase revenue by anywhere from .1% to 8%.

Question 2B: The randomization check ensures that the treatment group and control group are similar before the treatment is applied. If they are, then any observed differences must be due to the treatment effect. The analysis shows that there is no statistically significance difference between the two groups, meaning that the notion that there is no difference between the two is plausible.

Question 2C: The coefficient estimates imply that the presence of advertising increases revenue by roughly .7%. Since we can't reject the null hypothesis, it is plausible that advertising has no effect. However, because of the size of the standard error, advertising could still plausibly increase revenue by as much as 3%.

Question 2D: 1D looked at how revenue changed in treatment dmas after advertising was turned off in a particular month. Any observed change in revenue could be due to advertising, or could be due to other things that changed over time. The control group allows us to observe what expected revenue would have been in the treatment period, had advertising been maintained. It accounts for anything that could be changing over time, other than advertising

Question 2E: The R-Squared is 8.422e-06, but it doesn't matter because this is a randomized experiment.

The R-Squared does not affect the interpretation or confidence in our results. While a low R-Squared implies our control variables do not explain a large part of the variance, they need only explain the difference between the treatment and control group in the experiment. This is indicated by the standard error. As a whole, even important omitted variables do not bias the results of an A/B test.