Liqi Zhu's Homework 1

Part 1

d)

```
Call:
lm(formula = log(TreatGroup$revenue) ~ TreatGroup$isTreatmentPeriod)

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 ***
TreatGroup$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
```

e)

```
Call:
lm(formula = log(PreTreatPeriod$revenue) ~ PreTreatPeriod$isTreatmentGroup)

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 ***
PreTreatPeriod$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
```

f)

```
Call:
lm(formula = log(PostTreatPeriod$revenue) ~ PostTreatPeriod$isTreatmentGroup)

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 ***
PostTreatPeriod$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
```

g)

```
Call:
lm(formula = log(PostTreatPeriod$revenue) ~ PostTreatPeriod$isTreatmentGroup *
    month)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0148 -0.7560 -0.0299  0.7402  3.8431

Coefficients:
                                       Estimate Std. Error t value Pr(>|t|)
(Intercept)                            10.89370    0.04634 235.068   <2e-16 ***
PostTreatPeriod$isTreatmentGroup       -0.02269    0.05636  -0.403    0.687
month6                                  0.02905    0.05351   0.543    0.587
month7                                  0.02531    0.05589   0.453    0.651
PostTreatPeriod$isTreatmentGroup:month6  0.02013    0.06508   0.309    0.757
PostTreatPeriod$isTreatmentGroup:month7  0.01539    0.06797   0.226    0.821
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.208 on 13014 degrees of freedom
Multiple R-squared:  0.0001688, Adjusted R-squared:  -0.0002154
F-statistic: 0.4393 on 5 and 13014 DF,  p-value: 0.8213
```

Part 2

a)

> In part 1d, you ran the analysis without a control group. What do the resulting coefficient estimates say about the effectiveness of advertising? Be as specific as you can. The best answers to this question will quantify the effect.

$log(u) - log(v) = log(\frac{u}{v})$
Coefficient shows search ads leads to 3.94% (100xcoefficient) decrease in revenue, since P-value 0.0474 < 0.05 (rejection of null hypoethesis).

b)

> What is the purpose of the randomization check in part 1e? What do the results of this analysis show?

Randomization check's purpose is to ensure there's no difference between treatment group and control group. This analysis shows that revenue is not relavent with treatment, since P-value 0.57 > 0.05 and null hypoethesis cannot be rejected.

c)

> In part 1f, you ran the analysis with a control group. What do the resulting coefficient estimates say about the effectiveness of advertising? Be as specific as you can. The best answers to this question will quantify the effect.

Since P-value > 0.05 and null hypoethesis cannot be rejected, cofficient is 0.
Revenue is not relavent with treatment in pretreatment period.

d)

> What was the purpose of the control group here? What was unaccounted for in part 1d, but was accounted for in part 1f?

Control group's purpose is to control all the variables except treatment.
Revenue changes over time, i.e. time factor was unaccounted for in part 1d, but was accounted for in part 1f.

e)

> Using the summary function, note the R-Squared of the regression in part 1f in the pdf. Does this effect the interpretation or confidence in the estimate of the effectiveness of advertising?

R-Squared of regression in part 1f is small enough(-6.839e-05). So it won't effect the interpretation or confidence in estimate.

f)

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

It's the right choice since runing on log(revenue) shows the percentage thus decrease the variance to meet the assumption of equal variance. Simply using revenue may cause huge variance while error need to be drawn from the same distribution.