Assignment2

SIYU LIU

Question1

```
sample = 500000
t_c <-f %>% filter(f$Treatment == 1 & f$saw_ads == 1)
std <- sqrt(sd(t_c$sales)^2)
std
```

```
## [1] 5.965673
```

The standard deviation of sales in the treatment group and saw the ad is 5.965673

Question2a

```
lift <- (0 + 0.01) /0.5
lift
```

```
## [1] 0.02
```

The signal lift is 0.02

Question2b

```
power <- function(lift,c_ratio,sample){
  control = sample* c_ratio
  treatment  = sample *(1-c_ratio)
  ste = sqrt(std^2/control + std^2/treatment)
  signal = lift/std
  power = pwr.t2n.test(d = signal, n1 =control, n2 = treatment, sig.level = 0.05 )
  print(power$power)
}
```

```
power(lift,0.2,sample)
```

```
## [1] 0.1576508
```

```
power(lift,0.8,sample)
```

```
## [1] 0.1576508
```

```
power(lift,0.3,sample)
```

```
## [1] 0.192319
```

If 20% users are assigne to control group, the power will be 15.77%. If 80% users are assigne to control group, the power will be 15.77%. If 30% users are assigne to control group, the power will be 19.23%.

Question2c

```
optimize(power,c(0,1),tol = 0.01,maximum =T,lift = lift, sample = sample)
```

```
## [1] 0.2104251
## [1] 0.2104251
## [1] 0.1717337
## [1] 0.220096
## [1] 0.2200883
## [1] 0.2200883
## [1] 0.220096
```

```
## $maximum
## [1] 0.5
##
## $objective
## [1] 0.220096
```

When 50% sample are assigned to the control, the maxmized statistical power can be got.

Question3a

```r
budget = 2000
cost_per_person = function(budget, c_ratio,sample){
  cost = budget/(sample*(1-c_ratio))
  cost
}

cost1 = cost_per_person(budget, 0.2,sample)
cost2 = cost_per_person(budget, 0.3,sample)
cost3 = cost_per_person(budget, 0.8,sample)

cost1
```

```
## [1] 0.005
```

```r
cost2
```

```
## [1] 0.005714286
```

```r
cost3
```

```
## [1] 0.02
```

The average ad spend person is 0.005 when 20% sample are assigned in control group

The average ad spend person is 0.0057 when 30% sample are assigned in control group

The average ad spend person is 0.02 when 80% sample are assigned in control group

Question3b

```r
signal1 = cost1*2
signal2 = cost2*2
signal3  = cost3*2


signal1
```

```
## [1] 0.01
```

```r
signal2
```

```
## [1] 0.01142857
```

```r
signal3
```

```
## [1] 0.04
```

The signal lift is 0.01 when 20% sample are assigned in control group

The signal lift is 0.0114 when 30% sample are assigned in control group

The signal lift is 0.04 when 80% sample are assigned in control group

Question 3c

```r
power(signal1,0.2,sample)
```

```
## [1] 0.07612453
```

```r
power(signal2,0.3,sample)
```

```
## [1] 0.09518236
```

```r
power(signal3,0.8,sample)
```

```
## [1] 0.4747419
```

If 20% users are assigned to control group, the power will be 7.6%. If 30% users are assigned to control group, the power will be 9.5%. If 80% users are assigned to control group, the power will be 47.47%.

Question 3d The signals are different because of the cost changes. The increase in signal will increase the statistic power.

80% users assigned to control group with fixed budget of 2000 is the best among six options.

Question 4a

```r
signal4 = cost_per_person(budget, 0.2,sample)*2
signal5 = cost_per_person(budget, 0.3,sample)*1.9
signal6 = cost_per_person(budget, 0.8,sample)*1.2

signal4
```

```
## [1] 0.01
```

```r
signal5
```

```
## [1] 0.01085714
```

```r
signal6
```

```
## [1] 0.024
```

2Xcost is 0.01 1.9X cost is 0.0108 1.2X cost is 0.024

Question 4b

```r
power4 = power(signal4,0.2,sample)
```

```
## [1] 0.07612453
```

```r
power5 = power(signal5, 0.3,sample)
```

```
## [1] 0.09069718
```

```r
power6 = power(signal6,0.8,sample)
```

```
## [1] 0.2064888
```

```r
power4
```

```
## [1] 0.07612453
```

```r
power5
```

```
## [1] 0.09069718
```

```r
power6
```

```
## [1] 0.2064888
```

If 20% users are assigned to control group, the power will be 7.6%. If 30% users are assigned to control group, the power will be 9.1%. If 80% users are assigned to control group, the power will be 20.6%.

20% user assigne to control group with 2X cost is the best in this setting due to the biggger signal

Question 5a

```
sample1 = 60
budget1 = 40000
lift7 <- cost_per_person(budget1,0.5,sample1)*3
lift7
```

```
## [1] 4000
```

The signal is 4000

5b

```
std1 = 30000
power1 <- function(lift,c_ratio,std1,sample,sig){
  control = sample* c_ratio
  treatment  = sample *(1-c_ratio)
  ste = sqrt(std1^2/control + std1^2/treatment)
  signal = lift/std1
  power = pwr.t2n.test(d = signal, n1 =control, n2 = treatment, sig.level = sig )
  power
}

power7 <- power1(lift7, 0.5, 30000,sample1,0.1)
power7$power
```

```
## [1] 0.143871
```

The statistical power is 0.143871

5c

```
power1(lift7, 0.5,30000,sample1*11,0.1)
```

```
##
##      t test power calculation
##
##               n1 = 330
##               n2 = 330
##                d = 0.1333333
##        sig.level = 0.1
##            power = 0.5267398
##      alternative = two.sided
```

If the current mean and std is the same, we need to run this experiment for 11 weeks to get statistical power to exceed 50%

6a

```
sample_t <- 0.002*300000
sample_c <- 0.002*100000
sample_t
```

```
## [1] 600
```

```
sample_c
```

```
## [1] 200
```

I will expect 600 in treatment and 200 in countrol group.

6b

```r
lift8 <- 4.1*0.02
power1(lift8,0.25,0.68, 800,0.05)
```

```
##
##      t test power calculation
##
##              n1 = 200
##              n2 = 600
##               d = 0.1205882
##       sig.level = 0.05
##           power = 0.31419
##     alternative = two.sided
```

```r
lift9 <- 1.6*0.04
power1(lift9,0.25,1.23, 800, 0.05)$power
```

```
## [1] 0.09754985
```

There are respectively 31.4% and 9.7% probability that the brand favorability and intention estimates exclude 0

6c I will not use DynemicLogics'services. One of the obvious weekness of survey analysis is that response rate is too low hence it is very hard for us to get enough sample size to increase the statistical power. however, survey analysis also have some advantages in this case as we can combined the online campaign with the off-line sales.

6d

```r
sample_t1 <- 0.005*300000
sample_c1 <- 0.001*100000
sample_t1
```

```
## [1] 1500
```

```r
sample_c1
```

```
## [1] 100
```

```r
power1(lift8,1/16,0.68,1600,0.05)
```

```
##
##      t test power calculation
##
##              n1 = 100
##              n2 = 1500
##               d = 0.1205882
##       sig.level = 0.05
##           power = 0.2147505
##     alternative = two.sided
```

```r
power1(lift9,1/16,0.68,1600,0.05)
```

```
##
##      t test power calculation
##
```

```
##              n1 = 100
##              n2 = 1500
##               d = 0.09411765
##       sig.level = 0.05
##           power = 0.1490859
##     alternative = two.sided
```

There are respectively 21.4% and 14.9% probability that the brand favorability and intention estimates exclude 0

althouhg the sample size is increase, the less of control and group increase the standard error and lower the power.