BA860 Assignment 2

Qiaoling Huang (U20421641)

2/11/2020

I collaborated with Salina(Ziqin Ma) for the assignment 2.

```
#install.packages("pwr")
library(pwr)
library(data.table)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()
```

```
data = fread("Downloads/AdFX-BA860-SectionA-W20-4004106-rows.csv")
data$gender[data$gender == "female"] <- 1
data$gender[data$gender == "male"] <- 0
data$gender <- as.numeric(data$gender)
glimpse(data)
```

```
## Observations: 4,004,106
## Variables: 5
## $ Treatment  <int> 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, ...
## $ saw_ads    <int> 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, ...
## $ sales      <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.0...
## $ past_sales <dbl> 0.00, 97.67, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0....
## $ gender     <dbl> 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, ...
```

```
## split treatment and control group
treat <- data[Treatment == 1]
control <- data[Treatment == 0]
treat_exp <- treat[saw_ads == 1]
control_exp <- control[saw_ads == 1]
```

1: What is the standard deviation of sales (not past_sales) in the treatment group among exposed users who saw the campaign in “AdFX-Class-Term- Row_count.csv” (from Assignment 1)? Use this number for your calculations below.

```r
## calculate standard standard deviation
sd_sales <- sd(treat_exp$sales)
print(sd_sales)
```

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

Answer: The standard deviation is 5.965673.

2-a:If your profitability margin on sales is 50%, what is the 'reasonable signal' for AdFX lift that you are assuming?

```r
## ROI = 0
lift = 2*0.01*(1+0/100)
print(lift)
```

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

Answer: The lift('reasonable signal) is 0.02.

2-b: For a 95% confidence interval, calculate the statistical power to detect a successful campaign for the following three potential experimental designs

```r
## 20% of users are assigned to the control group
pwr.t2n.test(d=lift/sd_sales, n1=500000*(1-0.2), n2=500000*0.2, sig.level = 0.05)
```

```
##
##      t test power calculation
##
##              n1 = 4e+05
##              n2 = 1e+05
##               d = 0.003352514
##       sig.level = 0.05
##           power = 0.1576508
##     alternative = two.sided
```

```r
## 30% of users are assigned to the control group
pwr.t2n.test(d=lift/sd_sales, n1=500000*(1-0.3), n2=500000*0.3, sig.level = 0.05)
```

```
##
##      t test power calculation
##
##              n1 = 350000
##              n2 = 150000
##               d = 0.003352514
##       sig.level = 0.05
##           power = 0.192319
##     alternative = two.sided
```

```r
## 80% of users are assigned to the control group
pwr.t2n.test(d=lift/sd_sales, n1=500000*(1-0.8), n2=500000*0.8, sig.level = 0.05)
```

```
##
##      t test power calculation
##
##              n1 = 1e+05
##              n2 = 4e+05
##               d = 0.003352514
##       sig.level = 0.05
##           power = 0.1576508
##     alternative = two.sided
```

Answer: 20% Control power:0.1576508 30% Control power:0.192319 80% Control power:0.1576508 The second design with 30% in control group is better according to this criterion.

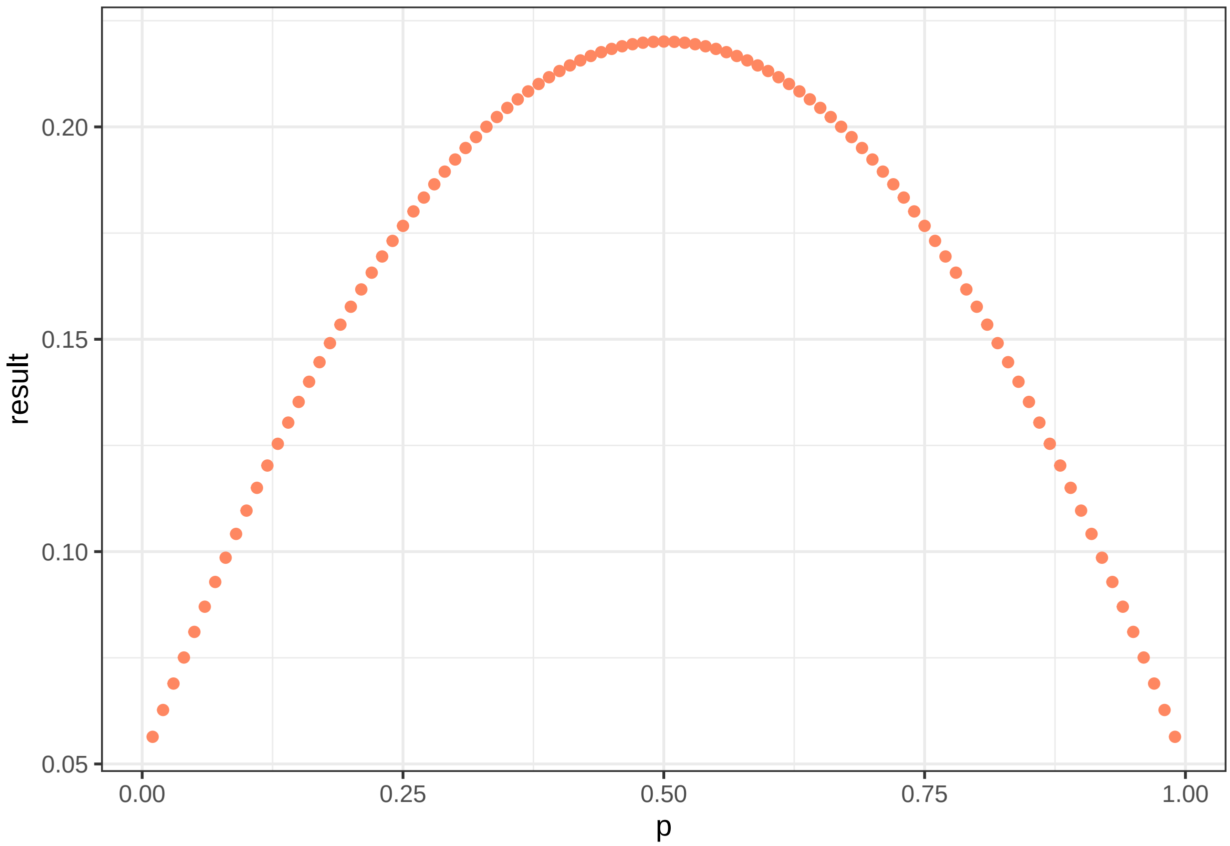
2_c:What treatment assignment (% assigned to the control group) maximizes statistical power?

```r
## create a power function
power_test<- function(p) {
  temp<-pwr.t2n.test(d=0.02/5.965673, n1=500000*p, n2=500000*(1-p), sig.level = 0.05)
  return(temp$power)
}

## create a seq p
p = seq(0.01,0.99,by=0.01)
p
```

```
##  [1] 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10 0.11 0.12 0.13 0.14
## [15] 0.15 0.16 0.17 0.18 0.19 0.20 0.21 0.22 0.23 0.24 0.25 0.26 0.27 0.28
## [29] 0.29 0.30 0.31 0.32 0.33 0.34 0.35 0.36 0.37 0.38 0.39 0.40 0.41 0.42
## [43] 0.43 0.44 0.45 0.46 0.47 0.48 0.49 0.50 0.51 0.52 0.53 0.54 0.55 0.56
## [57] 0.57 0.58 0.59 0.60 0.61 0.62 0.63 0.64 0.65 0.66 0.67 0.68 0.69 0.70
## [71] 0.71 0.72 0.73 0.74 0.75 0.76 0.77 0.78 0.79 0.80 0.81 0.82 0.83 0.84
## [85] 0.85 0.86 0.87 0.88 0.89 0.90 0.91 0.92 0.93 0.94 0.95 0.96 0.97 0.98
## [99] 0.99
```

```r
## apply p to the function
results = lapply(p, power_test)
result = unlist(results)
## put the result and p into a data frame for plot
result_tab = data.frame(p, result)
## plot the result
ggplot(result_tab, aes(p, result))+
  geom_point(colour = "#fe8761")+
  theme_bw()
```



Answer: The statistical power achieve around 50% when p equal 0.5.

3: Continuing from Q1 & Q2, Now suppose instead that you have a fixed ad budget of $2,000 to spend on your own ads. You ignore the cost of control ads because your partner the publisher is providing these for free. a: Calculate the average ad spend per person for the three experimental designs above.

```
## 20%
cost_per_person1 <- 2000/(500000* (1-0.2))
print(cost_per_person1)
```

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

```
## 30%
cost_per_person2 <- 2000/(500000* (1-0.3))
print(cost_per_person2)
```

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

```
## 80%
cost_per_person3 <- 2000/(500000* (1-0.8))
print(cost_per_person3)
```

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

Answer: 20% Control cost per person:0.005 30% Control cost per person:0.005714286 80% Control cost per person:0.02

b. Maintaining your assumption that the campaign will break-even for each experimental design, what is the 'reasonable signal' for each of the three possible experimental designs above?

```
lift_assumption1 <- 2*cost_per_person1*(1+0/100)
print(lift_assumption1)
```

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

```
lift_assumption2 <- 2*cost_per_person2*(1+0/100)
print(lift_assumption2)
```

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

```
lift_assumption3 <- 2*cost_per_person3*(1+0/100)
print(lift_assumption3)
```

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

Answer: 20% Control lift:0.01 30% Control lift:0.01142857 80% Control lift:0.04

c: For a 95% confidence interval, what is the statistical power now for each of the three possible experimental designs? Which is highest now?

```
## 20% of users are assigned to the control group
pwr.t2n.test(d=lift_assumption1/sd_sales, n1=500000*(1-0.2), n2=500000*0.2, sig.level = 0.05)
```

```
##
##      t test power calculation
##
##             n1 = 4e+05
##             n2 = 1e+05
##              d = 0.001676257
##      sig.level = 0.05
##          power = 0.07612453
##    alternative = two.sided
```

```
## 30% of users are assigned to the control group
pwr.t2n.test(d=lift_assumption2/sd_sales, n1=500000*(1-0.3), n2=500000*0.3, sig.level = 0.05)
```

```
##
##      t test power calculation
##
##             n1 = 350000
##             n2 = 150000
##              d = 0.001915722
##      sig.level = 0.05
##          power = 0.09518236
##    alternative = two.sided
```

```r
## 80% of users are assigned to the control group
pwr.t2n.test(d=lift_assumption3/sd_sales, n1=500000*(1-0.8), n2=500000*0.8, sig.level = 0.05)
```

```
##
##      t test power calculation
##
##              n1 = 1e+05
##              n2 = 4e+05
##               d = 0.006705027
##       sig.level = 0.05
##           power = 0.4747419
##     alternative = two.sided
```

Answer: 20% Control Power:0.07612453 30% Control Power:0.09518236 80% Control Power:0.4747419 The third experiment with 80% in control group is the highest now.

d: What changes your answer between Q2b) and Q3c)? Which is the best of the six options in terms of statistical power? Answer: The cost spent per person is higher and the statiscal power is higher.

4_a.What is the 'reasonable' signal now? Fill out the above table.

```r
## reasonable signal
resig1<-2*cost_per_person1
print(resig1)
```

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

```r
resig2<-1.9*cost_per_person2
print(resig2)
```

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

```r
resig3<-1.2*cost_per_person3
print(resig3)
```

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

Answer: 20% Control lift:0.01 30% Control lift:0.01085714 80% Control lift:0.024

4_b:Under your revised assumption, what is the statistical power of each option? Which is best now?

```r
## 20% of users are assigned to the control group
pwr.t2n.test(d=resig1/sd_sales, n1=500000*(1-0.2), n2=500000*0.2, sig.level=0.05)
```

```
##
##      t test power calculation
##
##              n1 = 4e+05
##              n2 = 1e+05
##               d = 0.001676257
##       sig.level = 0.05
##           power = 0.07612453
##     alternative = two.sided
```

```r
## 30% of users are assigned to the control group
pwr.t2n.test(d=resig2/sd_sales, n1=500000*(1-0.3), n2=500000*0.3, sig.level=0.05)
```

```
##
##      t test power calculation
##
##              n1 = 350000
##              n2 = 150000
##               d = 0.001819936
##       sig.level = 0.05
##           power = 0.09069718
##     alternative = two.sided
```

```r
## 80% of users are assigned to the control group
pwr.t2n.test(d=resig3/sd_sales, n1=500000*(1-0.8), n2=500000*0.8, sig.level=0.05)
```

```
##
##      t test power calculation
##
##              n1 = 1e+05
##              n2 = 4e+05
##               d = 0.004023016
##       sig.level = 0.05
##           power = 0.2064888
##     alternative = two.sided
```

Answer: 20% Control Power:0.07612453 30% Control Power:0.09069718 80% Control Power:0.2064888 The last experiment with 80% users in control group has the best statistical power as opposed to others.

5-a:You budget $40,000 weekly for the paid search campaign. For your power calculation, you expect the search ads to generate a lift equivalent to 3X the ad cost. What is the expected ad lift for the average store (the 'reasonable' signal)?

```r
## cost per mda in budget 40,000
cost_per_mda <- 40000/(60/2)
cost_per_mda
```

```
## [1] 1333.333
```

```r
avg_lift_mda <- 3*cost_per_mda

print(avg_lift_mda)
```

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

Answer: The expect ad lift for the average store is 4000.

5_b:Across the 60 locations, the stores have the same average weekly sales of $200,000 with standard deviation $30,000. Given this and using a 90% confidence interval as your standard, calculate the statistical power for this experiment if you run it for a single week.

```r
pwr.t2n.test(d = avg_lift_mda/30000, n1=30, n2=30, sig.level=0.1)
```

```
##
##      t test power calculation
##
##              n1 = 30
##              n2 = 30
##               d = 0.1333333
##       sig.level = 0.1
##           power = 0.143871
##     alternative = two.sided
```

Answer: The statistical power is 0.143871 if we run it for a single week.

5_c: At current spending levels, how many weeks would you need to run this experiment before the statistical power of the experiment surpasses 50%?

```r
power_0.5<-pwr.t2n.test(power = 0.5, n1=30, n2=30, sig.level=0.1)
d=power_0.5$d
d
```

```
## [1] 0.4293812
```

```r
## d = avg_lift*weeks/sd
weeks = d*30000/avg_lift_mda
weeks
```

```
## [1] 3.220359
```

Answer: We would need to run this experiment at least 3 weeks.

6-a: Given the response rates, how many surveys do you expect to collect in each of the treatment and control groups?

```r
treat_survey = 300000*0.002
treat_survey
```

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

```r
control_survey = 100000*0.002
control_survey
```

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

Answer: We expect to collect 600 surveys from treatment group and 200 surverys from control group.

6_b:DynamicLogic survey proposal is expensive and you can only pay them for this one ad campaign. What is the likelihood that the 95% confidence intervals on your AdFX estimates exclude 0 for each survey measure?

```r
## Baseline brand favorability (4.1 out of 5)
favorability_baseline = 0.02*4.1
sd_baseline = 0.68
## Intent to purchase measure (1.6 out of 5)
favorability_intend = 0.04 * 1.6
sd_intend = 1.23

## statistical power favorability
pwr.t2n.test(d = favorability_baseline/sd_baseline, n1 = 300000*0.002, n2 = 100000*(0.002), sig.level =
```

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

```r
## statistical power favorability 2
pwr.t2n.test(d = favorability_intend/sd_intend, n1 = 300000*0.002, n2 = 100000*(0.002), sig.level = 0.05
```

```
##
##      t test power calculation
##
##              n1 = 600
##              n2 = 200
##               d = 0.05203252
##       sig.level = 0.05
##           power = 0.09754985
##     alternative = two.sided
```

Answer: The statistical power for the favorability of brand baseline is 0.31419, and the statistical power for the facorability of intend to purchase is 0.09754985. The results show it's unlikely that the 95% condidence intervals on AdFX estimates exclude 0 for each survey measure.

6_c: In a one-paragraph summary for a coworker, explain your decision on whether to use DynamicLogic's services. Be sure to discuss the strengths and limitations of survey analysis and explain implications of your power calculation for the experiment.

Answer: Strengths of surveys: 1: It has high representativeness which represents a large populations. 2: Low costs. 3: Easy to gather data. 4: It's easy to find statistically significant results other others data gathering methods. 5: High reliability and usually researcher's own bias are eliminated. 6: Precise results.

Limitations of surverys: 1: Inflexible design, meaning not be able to change the survery throughout the process. 2: Not ideal for controversial issues, not be able to face-to-face interviews. 3: The questions created from the surveys may not be appropriate for all the participants.

Reference:https://explorable.com/advantages-and-disadvantages-of-surveys?from=singlemessage&isappinstalled=0

I will not use DynamicLogic's services, as the staticical power of both survery measures are less than 0.5 meaning the results are not statistically significant. Also, surveys do not cost a lot of money to create them, DynamicLogic's services could've sent more surveys to users if there is only 0.2% people expect to fill out the surveys. It can help to increase the statistical power if there are more people fill out the surveys and the results will be more signicant.

6_d: Suppose that 0.5% of users who see the treatment ad fill out the survey, but 0.1% of users who see the control ad fill the survey. How would this affect your interpretation of the experimental results?

```
pwr.t2n.test(d = favorability_baseline/sd_baseline, n1 = 300000*0.005, n2 = 100000*(0.001), sig.level =
```

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

```
pwr.t2n.test(d = favorability_intend/sd_intend, n1 = 300000*0.005, n2 = 100000*(0.001), sig.level = 0.05
```

```
##
##      t test power calculation
##
##              n1 = 1500
##              n2 = 100
##               d = 0.05203252
##       sig.level = 0.05
##           power = 0.07951229
##     alternative = two.sided
```

Answer: The statistical power for both survey measures are 0.2147505 and 0.07951229 which are even smaller than before. We think the gap between the number of people who fill out the surveys in treatment group and control is too large. And the number of people who fill out the surveys in control group is way too small, which cannot truely represent the entire group population and the standard deviation can be very large.