Online Campaign Evaluation

Ad Effectiveness Experiments

2/6/2020

Pairs: Sherry Zuo, Yishuang Song

In the Yahoo! retail experiment case, a manager evaluates the effectiveness of a retailer's ad campaign using an experiment. In this assignment, you are the marketing manager at Nordsaksingdale's, an apparel retailer. You ran an experiment to measure the effectiveness of a recent ad campaign and must analyze the resulting dataset: "AdFXClass-Term-Row_count.csv" (Canvas).

```
#load library
library(readr)
library(tidyverse)
```

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

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v ggplot2 3.2.1     v forcats 0.4.0
```

```
## -- Conflicts ------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#load data
AdFx<-read_csv("AdFX-BA860-SectionB-W20-4004106-rows.csv")
```

```
## Parsed with column specification:
## cols(
##   Treatment = col_double(),
##   saw_ads = col_double(),
##   sales = col_double(),
##   past_sales = col_double(),
##   gender = col_character()
## )
```

```
dim(AdFx)
```

```
## [1] 4004106       5
```

```
names(AdFx)
```

```
## [1] "Treatment"  "saw_ads"    "sales"      "past_sales" "gender"
```

Treatment Equals 1 if user is in Treatment group & 0 if user is in Control group
Sales Sales at the retailer in 2 weeks of campaign ()$Sales_past Sales at the retailer in 2 weeks prior to the campaign()$
Gender Male or Female user segments
Saw_ads Equals 1 if user saw 1+ experimental ad (retailer ad for users in Treatment group and control ad for user in Control group) & 0 otherwise

1. Before analyzing the experiment's results, we want to verify that the experiment properly randomized users. Otherwise, we will not be very confident in our results.
   To do this, we compare the treatment and control groups by the users' baseline characteristics.

a. (15 pts) Verify the randomization by user gender

b. (10 pts) Use the experimental differences estimator to compare the proportion of women in the Treatment versus Control groups(hint: convert the data to numeric using 1=female & 0=male to compute the difference). Report the absolute and relative differences, s.e. and c.i..

```
AdFx=AdFx %>%
  mutate(gender = ifelse(gender=="female", 1, 0))
```

```
Treatment_group<-AdFx%>%filter(Treatment==1)
Control_group<-AdFx%>%filter(Treatment==0)
```

```
absolute_difference_w<-mean(Treatment_group$gender)-mean(Control_group$gender)
absolute_difference_w
```

```
## [1] 0.0007840679
```

```
relative_difference_w<-
  (mean(Treatment_group$gender)-mean(Control_group$gender))*100/mean(Control_group$gender)
relative_difference_w
```

```
## [1] 0.152154
```

```
diffSE_w <- sqrt(sd(Treatment_group$gender)^2 / length(Treatment_group$gender)
            + sd(Control_group$gender)^2 / length(Control_group$gender))
diffSE_w
```

```
## [1] 0.0005237205
```

```
ciLow_w <- absolute_difference_w - 1.96*diffSE_w
ciHigh_w <- absolute_difference_w + 1.96*diffSE_w
ciLow_w
```

```
## [1] -0.0002424242
```

```
ciHigh_w
```

```
## [1] 0.00181056
```

Answer: The absolute difference is **0.0007840679**, the relative difference is **0.152154%**, the standard error is **0.0005237205**, the confidence interval is [**-0.0002424242, 0.00181056**]

ii. (5 pts) What do you conclude about the validity of the experimental randomization in terms of gender? *Answer:* We conclude that the experimental randomization in terms of gender is **valid** because the difference is almost equal to 0 and 0 is not excluded in the confidence interval, so the difference is not statistically significant, so the proportion of women in treatment and control is the same, and it's randomized.

b. (15 pts) Verify the randomization by past sales

c. (10 pts) Use the experimental differences estimator to compare the average sales in the 2 weeks before the experiment in the Treatment versus Control groups. Report the absolute and relative differences, s.e. and c.i..

```
absolute_difference_ps<-mean(Treatment_group$past_sales)-mean(Control_group$past_sales)
absolute_difference_ps
```

```
## [1] 0.00518405
```

```
relative_difference_ps<-
  (mean(Treatment_group$past_sales)-mean(Control_group$past_sales))*100/mean(Control_group$past_sales)
relative_difference_ps
```

```
## [1] 0.436697
```

```
diffSE_ps <- sqrt(sd(Treatment_group$past_sales)^2 / length(Treatment_group$past_sales)
           + sd(Control_group$past_sales)^2 / length(Control_group$past_sales))
diffSE_ps
```

```
## [1] 0.009019051
```

```
ciLow_ps <- absolute_difference_ps - 1.96*diffSE_ps
ciHigh_ps <- absolute_difference_ps + 1.96*diffSE_ps
ciLow_ps
```

```
## [1] -0.01249329
```

```
ciHigh_ps
```

```
## [1] 0.02286139
```

Answer: The absolute difference is **0.00518405**, the relative difference is **0.436697%**, the standard error is **0.009019051**, the confidence interval is [**-0.01249329, 0.02286139**]

ii. (5 pts) What do you conclude about the validity of the experimental randomization in terms of past sales?

Answer: We conclude that the experimental randomization in terms of past sales is **valid** because the difference is almost equal to 0 and 0 is not excluded in the confidence interval, so the difference is not statistically significant, so the past sales in treatment and control is the same, and it's randomized.

2. What would you estimate to be the effect of the campaign using an experiment that did not have control ads? Compute the experimental estimate for all users in the experiment: the (average) Intention to Treat estimate. Report the absolute and relative differences, s.e. and c.i..

```
absolute_difference_s<-mean(Treatment_group$sales)-mean(Control_group$sales)
absolute_difference_s
```

```
## [1] 0.02419101
```

```
relative_difference_s<-
  (mean(Treatment_group$sales)-mean(Control_group$sales))*100/mean(Control_group$sales)
relative_difference_s
```

```
## [1] 2.215343
```

```
diffSE_s <- sqrt(sd(Treatment_group$sales)^2 / length(Treatment_group$sales)
           + sd(Control_group$sales)^2 / length(Control_group$sales))
diffSE_s
```

```
## [1] 0.006583232
```

```
ciLow_s <- absolute_difference_s - 1.96*diffSE_s
ciHigh_s <- absolute_difference_s + 1.96*diffSE_s
ciLow_s
```

```
## [1] 0.01128787
```

```
ciHigh_s
```

```
## [1] 0.03709414
```

Answer: The absolute difference is **0.02419101**, the relative difference is **2.215343%**, the standard error is **0.006583232**, the confidence interval is [**0.01128787, 0.03709414**]

3. This experiment used control ads. Verify that the control ads were deployed the same as the retailer ads by comparing the Treatment and Control groups among the subset of exposed users.

a. (15 pts) Verify the equivalence of Treatment exposed and Control exposed users by gender (repeat both steps in question 1A).

```
Treatment_exposed<-Treatment_group%>%
  filter(saw_ads==1)
Control_exposed<-Control_group%>%
  filter(saw_ads==1)
absolute_difference_we<-mean(Treatment_exposed$gender)-mean(Control_exposed$gender)
absolute_difference_we
```

```
## [1] -0.1063946
```

```r
relative_difference_we<-
  (mean(Treatment_exposed$gender)-mean(Control_exposed$gender))*100/mean(Control_exposed$gender)
relative_difference_we
```

```
## [1] -17.08321
```

```r
diffSE_we <- sqrt(sd(Treatment_exposed$gender)^2 / length(Treatment_exposed$gender)
           + sd(Control_exposed$gender)^2 / length(Control_exposed$gender))
diffSE_we
```

```
## [1] 0.00077941
```

```r
ciLow_we <- absolute_difference_we - 1.96*diffSE_we
ciHigh_we <- absolute_difference_we + 1.96*diffSE_we
ciLow_we
```

```
## [1] -0.1079223
```

```r
ciHigh_we
```

```
## [1] -0.104867
```

Answer: The absolute difference is **-0.1063946**, the relative difference is **-17.08321%**, the standard error is **0.00077941**, the confidence interval is [**-0.1079223, -0.104867**]. We conclude that the control ads in terms of gender are **not valid** because the difference is less than 0 and 0 is excluded in the confidence interval. Therefore, the difference is statistically significant, and the control ads were deployed not the same as the retailer ads by comparing the Treatment and Control groups among the subset of exposed users by gender, and the control ads are not valid.

b. (15 pts) Verify the equivalence of Treatment exposed and Control exposed users by past sales (repeat both steps in question 1B).

```r
absolute_difference_eps<-mean(Treatment_exposed$past_sales)-mean(Control_exposed$past_sales)
absolute_difference_eps
```

```
## [1] 0.1465299
```

```r
relative_difference_eps<-
  (mean(Treatment_exposed$past_sales)-mean(Control_exposed$past_sales))*100/
  mean(Control_exposed$past_sales)
relative_difference_eps
```

```
## [1] 14.1546
```

```r
diffSE_eps <- sqrt(sd(Treatment_exposed$past_sales)^2 / length(Treatment_exposed$past_sales)
           + sd(Control_exposed$past_sales)^2 / length(Control_exposed$past_sales))
diffSE_eps
```

```
## [1] 0.0149949
```

```r
ciLow_eps <- absolute_difference_eps - 1.96*diffSE_eps
ciHigh_eps <- absolute_difference_eps + 1.96*diffSE_eps
ciLow_eps
```

```
## [1] 0.1171399
```

```r
ciHigh_eps
```

```
## [1] 0.1759199
```

Answer: The absolute difference is **0.1465299**, the relative difference is **14.1546%**, the standard error is **0.0149949**, the confidence interval is [**0.1171399, 0.1759199**]. We conclude that the control ads in terms of past sales are **not valid** because the difference is larger than 0 and 0 is excluded in the confidence interval. Therefore, the difference is statistically significant, and the control ads were deployed not the same as the retailer ads by comparing the Treatment and Control groups among the subset of exposed users by past sales, and the control ads are not valid.

4. (10 pts) How does your ad effectiveness estimate change when you make use of the control ads? Compute the experimental estimate for those users who saw ads:
   the (average) Treatment on Treated (TOT) estimate. Report the absolute and relative differences, s.e. and c.i..

```r
absolute_difference_es<-mean(Treatment_exposed$sales)-mean(Control_exposed$sales)
absolute_difference_es
```

```
## [1] 0.1567826
```

```r
relative_difference_es<-
  (mean(Treatment_exposed$sales)-mean(Control_exposed$sales))*100/mean(Control_exposed$sales)
relative_difference_es
```

```
## [1] 19.05238
```

```r
diffSE_es <- sqrt(sd(Treatment_exposed$sales)^2 / length(Treatment_exposed$sales)
           + sd(Control_exposed$sales)^2 / length(Control_exposed$sales))
diffSE_es
```

```
## [1] 0.008682931
```

```r
ciLow_es <- absolute_difference_es - 1.96*diffSE_es
ciHigh_es <- absolute_difference_es + 1.96*diffSE_es
ciLow_es
```

```
## [1] 0.1397641
```

```r
ciHigh_es
```

```
## [1] 0.1738011
```

Answer: The absolute difference is **0.1567826**, the relative difference is **19.05238%**, the standard error is **0.008682931**, the confidence interval is [**0.1397641, 0.1738011**]

5. (5 pts each) What is the total effect of the campaign on sales? (reminder: still include standard error, etc.)

a. Compute the total effect using the ITT estimate. Report the absolute and relative differences, s.e. and c.i..

```
absolute_difference_ts<-absolute_difference_s*length(Treatment_group$Treatment)
absolute_difference_ts
```

```
## [1] 62989.61
```

```
relative_difference_ts<-relative_difference_s
relative_difference_ts
```

```
## [1] 2.215343
```

```
diffSE_ts<-diffSE_s*length(Treatment_group$Treatment)
diffSE_ts
```

```
## [1] 17141.71
```

```
ciLow_ts <- absolute_difference_ts - 1.96*diffSE_ts
ciHigh_ts <- absolute_difference_ts + 1.96*diffSE_ts
ciLow_ts
```

```
## [1] 29391.86
```

```
ciHigh_ts
```

```
## [1] 96587.36
```

Answer: Using ITT estimate the total effect, the difference is **62989.61**, the relative difference is **2.215343%**, the standard error is **17141.71**, the confidence interval is [**29391.86, 96587.36**]

b. Compute the total effect using the TOT estimate. Report the absolute and relative differences, s.e. and c.i..

```
absolute_difference_tes<-absolute_difference_es*length(Treatment_exposed$Treatment)
absolute_difference_tes
```

```
## [1] 177077
```

```
relative_difference_tes<-relative_difference_es
relative_difference_tes
```

```
## [1] 19.05238
```

```r
diffSE_tes<-diffSE_es*length(Treatment_exposed$Treatment)
diffSE_tes
```

```
## [1] 9806.876
```

```r
ciLow_tes <- absolute_difference_tes - 1.96*diffSE_tes
ciHigh_tes <- absolute_difference_tes + 1.96*diffSE_tes
ciLow_tes
```

```
## [1] 157855.5
```

```r
ciHigh_tes
```

```
## [1] 196298.5
```

Answer: Using TOT to estimate the total effect, the absolute difference is **177077**, the relative difference is **19.05238%**, the standard error is **9806.876**, the confidence interval of total sales is [**157855.5, 196298.5**].

c. Based on your analysis in question 3, which of the two estimates should you report from this experiment and why?

Answer: The control ads were deployed not the same as the retailer ads by comparing the Treatment and Control groups among the subset of exposed users, they have bias on on treated-exposed group and controled-exposed group, therefore we should report **ITT estimate** from this experiment.

d. Using your preferred estimator, summarize your results for a manager. What are the managerial and statistical implications of your results?

Answer: Dear manager, we think we should use the ITT estimator since the control ads are not validate, so we should only compare the total treatment and control group, our experiment design achieve randomization, the ad effect is valid since their confidence intervals exclude 0, the ad effect for sales is 0.02419101 dollars, and the total ad effect for sales is 62989.61 dollars. The total sales would in the range of 157855.5 dollars to 196298.5 dollars. Thus, we think we should launch the ad. However, since we do not have enough information for ROI, so there is no guarantee for our result.

6. (5 pts each) What would you estimate to be the effect of the campaign without an experiment?

a. Compute the observational estimate (exposed vs. unexposed in Treatment group). Report the absolute and relative differences, s.e. and c.i..

```r
Treatment_unexposed<-Treatment_group%>%
  filter(saw_ads==0)
absolute_difference_ets<-mean(Treatment_exposed$sales)-mean(Treatment_unexposed$sales)
absolute_difference_ets
```

```
## [1] -0.2410307
```

```r
relative_difference_ets<-
  (mean(Treatment_exposed$sales)-mean(Treatment_unexposed$sales))*100/mean(Treatment_unexposed$sales)
relative_difference_ets
```

```
## [1] -19.74502
```

```r
diffSE_ets <- sqrt(sd(Treatment_exposed$sales)^2 / length(Treatment_exposed$sales)
           + sd(Treatment_unexposed$sales)^2 / length(Treatment_unexposed$sales))
diffSE_ets
```

```
## [1] 0.007536113
```

```r
ciLow_ets <- absolute_difference_ets - 1.96*diffSE_ets
ciHigh_ets <- absolute_difference_ets + 1.96*diffSE_ets
ciLow_ets
```

```
## [1] -0.2558015
```

```r
ciHigh_ets
```

```
## [1] -0.2262599
```

Answer: The absolute difference is **-0.2410307**, the relative difference is **-19.74502%**, the standard error is **0.007536113**, the confidence interval is [**-0.2558015, -0.2262599**].

b. Suppose a manager had not run an experiment and only had the observational estimate. What would they get wrong?

Answer: People may not be seeing the ads for various reasons, and exposed users may differ from unexposed users for the observational estimate. Thus, the estimate sales different may be caused by those different users. This may causes selection bias, and could cause wrong results.

7. (5 pts each) Consider gender as a segmentation variable.

a. Using your preferred estimator from question 5C, what is the average ad effect for women? Report the absolute and relative differences, s.e. and c.i..

```r
Treatment_female<-Treatment_group%>%
  filter(gender==1)
Control_female<-Control_group%>%
  filter(gender==1)
absolute_difference_ws<-mean(Treatment_female$sales)-mean(Control_female$sales)
absolute_difference_ws
```

```
## [1] 0.03187513
```

```r
relative_difference_ws<-
  (mean(Treatment_female$sales)-mean(Control_female$sales))*100/mean(Control_female$sales)
relative_difference_ws
```

```
## [1] 2.306178
```

```r
diffSE_ws <- sqrt(sd(Treatment_female$sales)^2 / length(Treatment_female$sales)
           + sd(Control_female$sales)^2 / length(Control_female$sales))
diffSE_ws
```

```
## [1] 0.01099246
```

```r
ciLow_ws <- absolute_difference_ws - 1.96*diffSE_ws
ciHigh_ws <- absolute_difference_ws + 1.96*diffSE_ws
ciLow_ws
```

```
## [1] 0.01032991
```

```r
ciHigh_ws
```

```
## [1] 0.05342035
```

Answer: Since ITT estimator is better. For women, the absolute difference is **0.03187513**, the relative difference is **2.306178%**, the standard error is **0.01099246**, the confidence interval is [**0.01032991, 0.05342035**].

b. Using your preferred estimator from question 5C, what is the average ad effect for men? Report the absolute and relative differences, s.e. and c.i..

```r
Treatment_male<-Treatment_group%>%
  filter(gender==0)
Control_male<-Control_group%>%
  filter(gender==0)
absolute_difference_ms<-mean(Treatment_male$sales)-mean(Control_male$sales)
absolute_difference_ms
```

```
## [1] 0.0150256
```

```r
relative_difference_ms<-
  (mean(Treatment_male$sales)-mean(Control_male$sales))*100/mean(Control_male$sales)
relative_difference_ms
```

```
## [1] 1.917867
```

```r
diffSE_ms <- sqrt(sd(Treatment_male$sales)^2 / length(Treatment_male$sales)
           + sd(Control_male$sales)^2 / length(Control_male$sales))
diffSE_ms
```

```
## [1] 0.006883141
```

```r
ciLow_ms <- absolute_difference_ms - 1.96*diffSE_ms
ciHigh_ms <- absolute_difference_ms + 1.96*diffSE_ms
ciLow_ms
```

```
## [1] 0.001534645
```

```r
ciHigh_ms
```

```
## [1] 0.02851656
```

Answer: Since ITT estimator is better. For men, the absolute difference is **0.0150256**, the relative difference is **1.917867%**, the standard error is **0.006883141**, the confidence interval is [**0.001534645, 0.02851656**].

c. Summarize the managerial and statistical implications of your results for a manager who needs to decide how to allocate the ad budget across gender. How would you recommend allocating the budget?

Answer: For both female and male groups, the confidence interval excluded 0, so the ads are valid for both female and male groups. Also, we could find the ads' effect is higher for females than males(0.032>0.015). Thus, from a statistical perspective, we recommend that the company allocate more ad budgets for females. However, the female is the targeting of more companies in the real-life, so the cost may be more expensive than men. While there are still a lot of factors affecting ads effect and we need more information to do calculations and finally suggest which gender the company should pay more attention to.