SmartAd AB Testing

Richard King

2023-09-17

Introduction

Since I've become fairly proficient in R, I decided to start leaning more heavily into the statistical aspect of the language. I recently took some courses in statistics and statistical programming in R, and I wanted to do some projects to apply what I learned.

I found this data from an A/B ad test on Kaggle and decided to use it to practice running a two sample t-test. Participants in the test were divided into a control and exposed group and shown either dummy ad or a creative, interactive SmartAd brand ad. Response rates were recorded along with info on when participants responded and what device, browser, and operating system they were using.

I'm definitely still wrapping my head around some of these statistical concepts, so any feedback that will help me learn is greatly appreciated!

Data Exploration

We will begin by loading our data and doing a bit of exploration to better understand the data from this A/B test.

```r
library(tidyverse)
df <- read_csv("~/GitHub/SmartAd-AB-Testing/SmartAd AB Testing Data.csv")
glimpse(df)
```

```
## Rows: 8,077
## Columns: 9
## $ auction_id  <chr> "0008ef63-77a7-448b-bd1e-075f42c55e39", "000eabc5-17ce-413~
## $ experiment  <chr> "exposed", "exposed", "exposed", "control", "control", "co~
## $ date        <date> 2020-07-10, 2020-07-07, 2020-07-05, 2020-07-03, 2020-07-0~
## $ hour        <dbl> 8, 10, 2, 15, 15, 15, 15, 5, 0, 14, 13, 6, 8, 8, 15, 16, 1~
## $ device_make <chr> "Generic Smartphone", "Generic Smartphone", "E5823", "Sams~
## $ platform_os <dbl> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6~
## $ browser     <chr> "Chrome Mobile", "Chrome Mobile", "Chrome Mobile WebView",~
## $ yes         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0~
## $ no          <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

We should first confirm that all the values in the "auction_id" column are actually unique.

```r
sum(duplicated(df$auction_id))
```

```
## [1] 0
```

None of the "auction_id" values are duplicated. We should also check for any missing values that could skew the data.

```
sum(is.na(df))
```

```
## [1] 0
```

Great! None of our values are missing. Next let's look at how many and what percentage of users are in the control and exposed groups.

```
df %>%
  group_by(experiment) %>%
  summarise(count = n()) %>%
  mutate(percent = 100*count/sum(count))
```

```
## # A tibble: 2 x 3
##   experiment count percent
##   <chr>      <int>   <dbl>
## 1 control     4071    50.4
## 2 exposed     4006    49.6
```

The control and exposed groups appear to have approximately the same number of users.

We can make a quick plot to look better understand how many users responded each day that the A/B test ran for. We are specifically looking at users who responded to the ads, so we need to filter our results to those users who chose either "yes" or "no".

```
resp_by_date <- df %>%
  select(date, yes, no) %>%
  filter(yes == 1 | no == 1) %>%
  group_by(date) %>%
  summarize(yes = sum(yes), no = sum(no)) %>%
  pivot_longer(-date, names_to = "response", values_to = "count")

resp_by_date$response = factor(resp_by_date$response, levels = c("yes", "no")) # Reorder factor levels

date_plot <- ggplot(resp_by_date, aes(x = as.factor(date), y = count, fill = response)) +
  geom_bar(stat = "identity", position = "dodge") +
  ggtitle("User Responses Each Day") +
  labs(x = "Date", y = "Number of Responses") +
  scale_fill_discrete(name = "Responses", labels = c("Yes", "No")) +
  scale_x_discrete(labels = c("3 Jul", "4 Jul", "5 Jul", "6 Jul", "7 Jul", "8 Jul", "9 Jul", "10 Jul"))

date_plot
```

User Responses Each Day
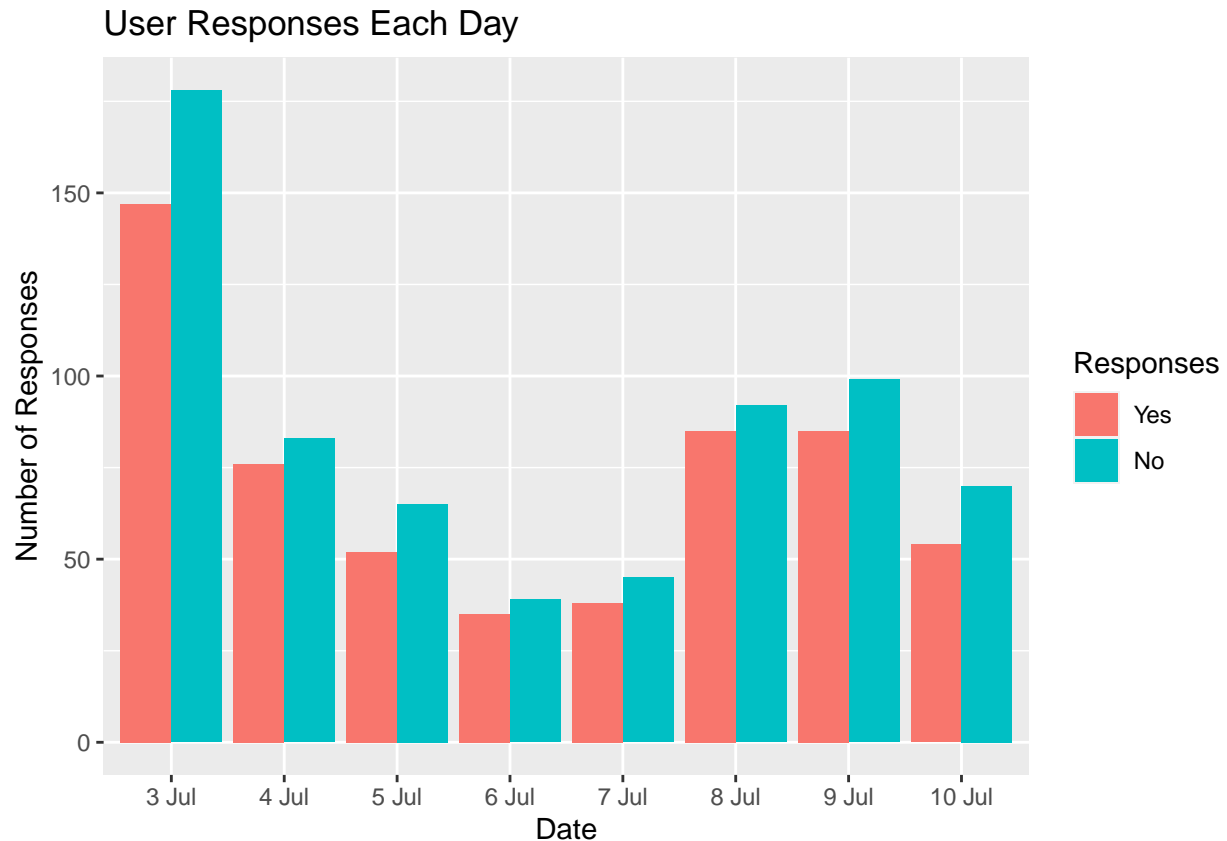
It seems a bit odd that the test would run eight days instead of just one week. 3 July 2020 was a Friday and had noticeably more responses than the other days. Let's also look at how user responses vary across an average day.

```
resp_by_hour <- df %>%
  select(hour, yes, no) %>%
  filter(yes == 1 | no == 1) %>%
  group_by(hour) %>%
  summarize(yes = sum(yes), no = sum(no)) %>%
  pivot_longer(-hour, names_to = "response", values_to = "count")

hour_plot <- ggplot(resp_by_hour, aes(x = as.factor(hour), y = count, fill = response)) +
  geom_bar(stat = "identity", position = "dodge") +
  ggtitle("User Responses Each Day") +
  labs(x = "Hour of the Day", y = "Number of Responses") +
  scale_fill_discrete(name = "Responses", labels = c("Yes", "No"))

hour_plot
```

User Responses Each Day
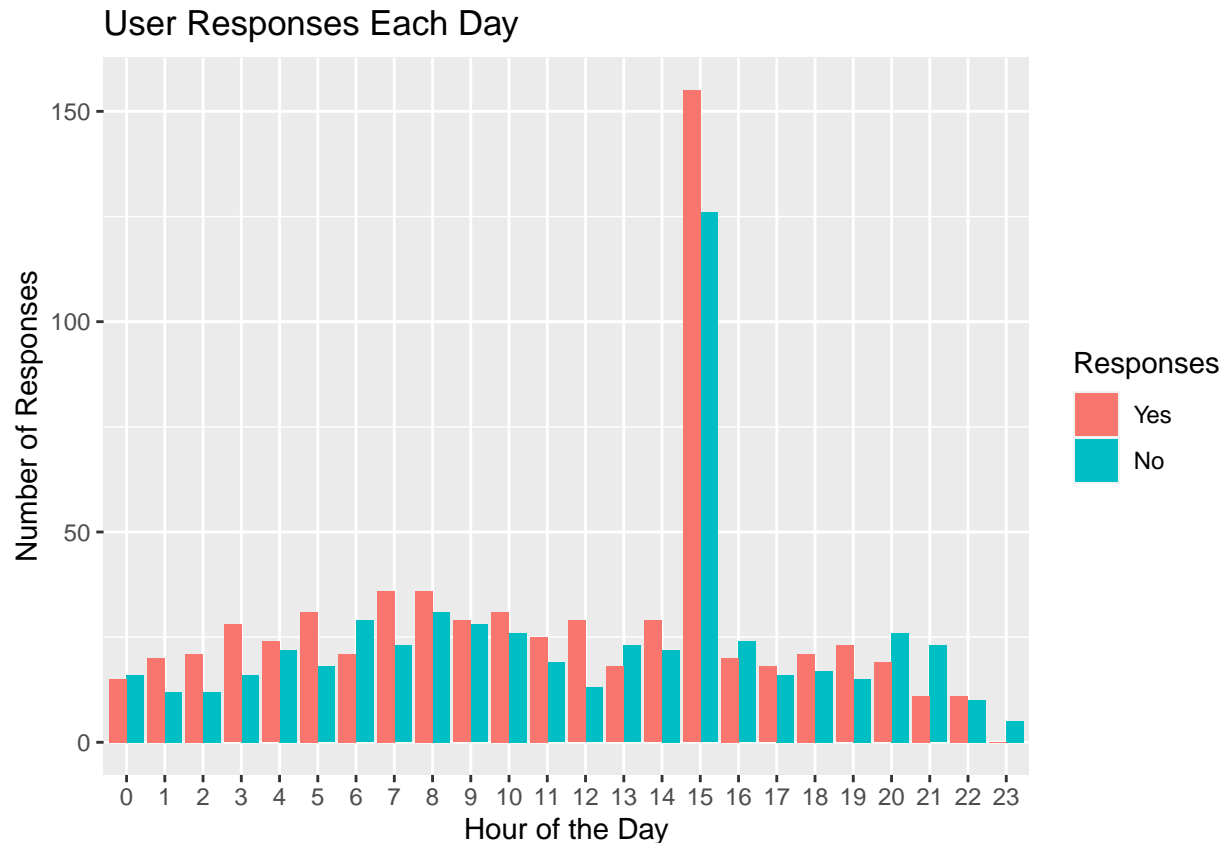
That is definitely interesting. I wonder what causes that spike at 3 pm.

Statistical Analysis

Now that we have gotten a better grasp on what our data looks like, we can move into our statistical analysis. We can use a two sample t-test to compare the mean response rates between our control and exposed groups. We first need to define our null and alternate hypotheses.

Null Hypothesis (H0): "There is no difference in response rates between the control and exposed groups."

Alternate Hypothesis (H1): "There is a difference in response rates between the control and exposed groups."

We will set our significance level (alpha) equal to 0.05. The response rate for the exposed group could be either higher or lower than that of the control group, so this will be a two-tailed test.

Since we are just interested in users from each group who responded "yes" or "no", we will create a data frame to capture just those users.

```
responses <- df %>%
  filter(yes == 1 | no == 1)
glimpse(responses)
```

```
## Rows: 1,243
## Columns: 9
## $ auction_id  <chr> "0016d14a-ae18-4a02-a204-6ba53b52f2ed", "008aafdf-deef-448~
## $ experiment  <chr> "exposed", "exposed", "exposed", "control", "control", "ex~
## $ date        <date> 2020-07-05, 2020-07-04, 2020-07-06, 2020-07-08, 2020-07-0~
```

```
## $ hour       <dbl> 2, 16, 8, 4, 15, 2, 15, 6, 15, 14, 15, 13, 20, 6, 20, 2, 1~
## $ device_make <chr> "E5823", "Generic Smartphone", "Generic Smartphone", "Sams~
## $ platform_os <dbl> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6~
## $ browser     <chr> "Chrome Mobile WebView", "Chrome Mobile", "Chrome Mobile",~
## $ yes         <dbl> 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1~
## $ no          <dbl> 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0~
```

It appears only 1,243 of the total 8,077 users responded to the ad. We should identify how many from each group responded "yes" or "no" as well as the conversion rate (the percent of users who responded "yes").

```
resp <- df %>%
  group_by(experiment) %>%
  summarise(yes = sum(yes), no = sum(no)) %>%
  mutate(count = yes + no) %>%
  mutate(conv_rate = yes/(yes + no)) %>%
  select(experiment, count, conv_rate, yes, no)
resp
```

```
## # A tibble: 2 x 5
##   experiment count conv_rate   yes    no
##   <chr>      <dbl>     <dbl> <dbl> <dbl>
## 1 control      586     0.451   264   322
## 2 exposed      657     0.469   308   349
```

Looks like the conversion rate for the control group is 45.1% and the rate for the exposed group is 46.9%. That seems like a pretty small difference, but the key question is whether the difference is statistically significant. Let's first reformat the responses for our control and exposed groups so that we can feed it into R's t.test() function.

```
control <- responses %>%
  mutate(control = ifelse(experiment == "control", yes, NA)) %>%
  select(control) %>%
  drop_na(control)

exposed <- responses %>%
  mutate(exposed = ifelse(experiment == "exposed", yes, NA)) %>%
  select(exposed) %>%
  drop_na(exposed)
```

Now we can run our t-test.

```
t.test(exposed, control, alternative = "two.sided", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  exposed and control
## t = 0.64537, df = 1225.7, p-value = 0.5188
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03730157  0.07387281
## sample estimates:
## mean of x mean of y
## 0.4687976 0.4505119
```

It appears our p-value is 0.5188, which is greater than our significance level of 0.05, which means we cannot reject our null hypothesis.

Conclusion

The fact that we cannot reject our null hypothesis suggests that there is no statistically significant difference in response rates between those who saw the SmartAd and those who saw the dummy ad. If the business running this test was considering switching to SmartAds, it would not be worth increasing spending to make that shift.

The peak in responses on 3 July and the spike in average responses at 3 pm each day are two curiosities that might be worth exploring in the future. I would also be interested in breaking down response rates by device, browser, and operating system and seeing if perhaps one of those variables causes a statistically significant difference in response rates.

Thanks for reading! As mentioned previously, feedback is always appreciated!