AB Test

Weifeng Wu

5/26/2022

```
library(plyr)
library(ggplot2)
library(scales)
```

```
abtest <- read.csv("ads_test.csv")
head(abtest)
```

```
##   id user_id                   timestamp     group advertising_pid clk
## 1  1  851104 2021-01-21 22:11:48.556739   control     430576_1007   0
## 2  2  804228 2021-01-12 08:01:45.159739   control     430576_1007   0
## 3  3  661590 2021-01-11 16:55:06.154213 treatment     430575_1007   0
## 4  4  853541 2021-01-08 18:28:03.143765 treatment     430575_1007   0
## 5  5  864975 2021-01-21 01:52:26.210827   control     430576_1007   1
## 6  6  936923 2021-01-10 15:20:49.083499   control     430576_1007   0
```

```
is.null(abtest)
```

```
## [1] FALSE
```

```
ddply(abtest,
      c("group"),
      summarise,
      rate = sum(clk)/length(clk))
```

```
##       group      rate
## 1   control 0.1204776
## 2 treatment 0.1229701
```

```
#Chi-square test
chisq.test(abtest$group,abtest$clk)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  abtest$group and abtest$clk
## X-squared = 4.255, df = 1, p-value = 0.03913
```

Since p-value $= 0.03913 < 0.05$, we cannot reject null hypothesis, at a 0.05 significane level, we can conclude that Pid have a significant difference.

```
abtest[,'day'] = strftime(abtest[,'timestamp'],"%D")
summary <- ddply(abtest,
                 c("day","group"),
                 summarise,
                 num=length(day),
                 rate=sum(clk)/length(clk))
```

```
head(summary)
```

```
##        day     group  num      rate
## 1 01/02/21   control 2895 0.1260794
## 2 01/02/21 treatment 2888 0.1239612
## 3 01/03/21   control 6682 0.1135887
## 4 01/03/21 treatment 6712 0.1175507
## 5 01/04/21   control 6666 0.1218122
## 6 01/04/21 treatment 6618 0.1207313
```

```
ggplot(summary,aes(x=day,y=rate,col=group,group=factor(group),lty = group))+
         geom_line(lwd=1)+
   geom_point(size=4)+
   geom_line(aes(y=rate,col=group))+
  geom_text(aes(label = round(rate,3), vjust = 1.1, hjust = -0.5, angle = 45), show.legend = FALSE)+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```