Twitter A/B Testing

Rayan Roy

Twitter is a microblogging and social networking service in which users communicate both publicly and privately via messages called "tweets". The intention of a tweet is to communicate a thought or idea as concisely and efficiently as possible – in 140 characters or less. Often, these tweets are used for advertising.

The primary mandate of the *Ads Data Science* team at Twitter is to increase ad revenue. If a user interacts with an ad, they may do so in a variety of different ways. They could view it, like it, comment on it, retweet it, click on it and be taken to the advertiser's websites, or some combination of these. A user's interaction with an ad indirectly generates revenue for Twitter, where increased engagement is associated with increased revenue.

The team is currently experimenting to determine which type of ad (Video, Image, or Text) maximizes engagement. However, a user's engagement with an ad may be influenced by more than just the ad type. For instance, a user's pre-existing relationship with the advertiser; a user's engagement with the ad may differ if the user

- follows the advertiser's Twitter account
- is followed by the advertiser's Twitter account
- neither follows, nor is followed by, the advertiser's Twitter account

Likewise, a user's engagement with the ad may be influenced by whether the ad is endorsed in same way (i.e., liked) by someone they have a personal connection with. For example:

- the ad is liked by one or more of the user's followees
- the ad is liked by one or more of the user's followers
- the ad is not liked by any of the user's followers or followees

A 3×3 Latin square design may be used to investigate the effects of ad type on engagement while controlling for the user's relationship with advertiser (RWA) and personal connection influence (PCI). Five hundred users of each RWA-PCI combination were assigned to the appropriate block, and on each an engagement score (a continuous measure that quantifies a user's engagement with an ad) was recorded. The data are available in the file `twitter-lsd.csv`.

Note that due to the Latin Square design, a user's block determined the type of ad they were shown. The table below visualizes this design. Note that V, P, and T respectively correspond to video, photo, and text ads.

		RWA		
		None	Following	Followed By
PCI	None	V	P	T
	Liked by Follower	T	V	P
	Liked by Followee	P	T	V

(a) What is the metric of interest and what is the corresponding response variable?

The metric of interest is the engagement rate, while the corresponding response variable is the engagement score (a continuous measure that quantifies a user's engagement with an ad).

(b) Identify which of the factors is the design factor and which are the nuisance factors. For all factors, state their levels.

The design factor is the ad type which has three levels namely video, image, text ads. The two nuisance factors are:

Design Factor 1: User's relationship with advertiser (RWA)

-Levels: None, Following, Followed By

Design Factor 2: User's personal connection influence (PCI)

-Levels: Not liked by both Follower or Followee, Liked by Follower, Liked by Followee

(c) What are the experimental units?

The experimental units are the users of Twitter from each RWA-PCI combination.

(d) State the relevant linear predictor for a regression-based analysis of this experiment. Be sure to define your notation, but adhere to the convention that the design factor is represented by x's and β's while the nuisance factors are represented by z's, w's, γ's and δ's.

$$\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \delta_1 w_{i1} + \delta_2 w_{i2}$$

where

$x_{i1} = 1$ if unit i is in a condition where the ad type is image and 0 if it is not.

$x_{i2} = 1$ if unit i is in a condition where the ad type is text and 0 if it is not.

$z_{i1} = 1$ if unit i is in a block with ad liked by follower and 0 if it is not.

$z_{i2} = 1$ if unit i is in a block with ad liked by followee and 0 if it is not.

$w_{i1} = 1$ if unit i is in a block with user following the advertiser's Twitter account and 0 if they are not.

$w_{i2} = 1$ if unit i is in a block with user being followed by the advertiser's Twitter account and 0 if they are not.

(e) State and test the hypothesis concerning the significance of the design factor. For full points, state the null distribution, the value of the test statistic, and the p-value. Draw your conclusion (in the context of the problem) at the 5% significance level.

The hypothesis is:

$$H_0 : \beta_1 = \beta_2 = 0 \ vs. \ H_A : \beta_j \neq 0 \ for \ some \ j = 1, 2$$

We test by comparing the full model (from part (d)) with the reduced model that is achieved by removing the β_j terms from the full model.

Using the partial F-test, we get:

```
twitter <- read.csv("twitter-lsd.csv")
full <- lm(twitter$Engagement ~ twitter$Ad.Type + twitter$PCI + twitter$RWA)
red <- lm(twitter$Engagement ~ twitter$PCI + twitter$RWA)
f.test <- anova(red, full)
f.test
```

```
## Analysis of Variance Table
##
## Model 1: twitter$Engagement ~ twitter$PCI + twitter$RWA
## Model 2: twitter$Engagement ~ twitter$Ad.Type + twitter$PCI + twitter$RWA
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   4495 88.721
## 2   4493 68.135  2    20.586 678.76 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
test.stat <- f.test$F[2]
p.value <- f.test$`Pr(>F)`
cat(paste("Test Statistic: ", test.stat, "\nP-value: ", p.value[2]))
```

```
## Test Statistic:  678.758804704515
## P-value:  2.63687745026101e-258
```

Therefore we get:

Null distribution: $T \sim \ F_{(2,4493)}$

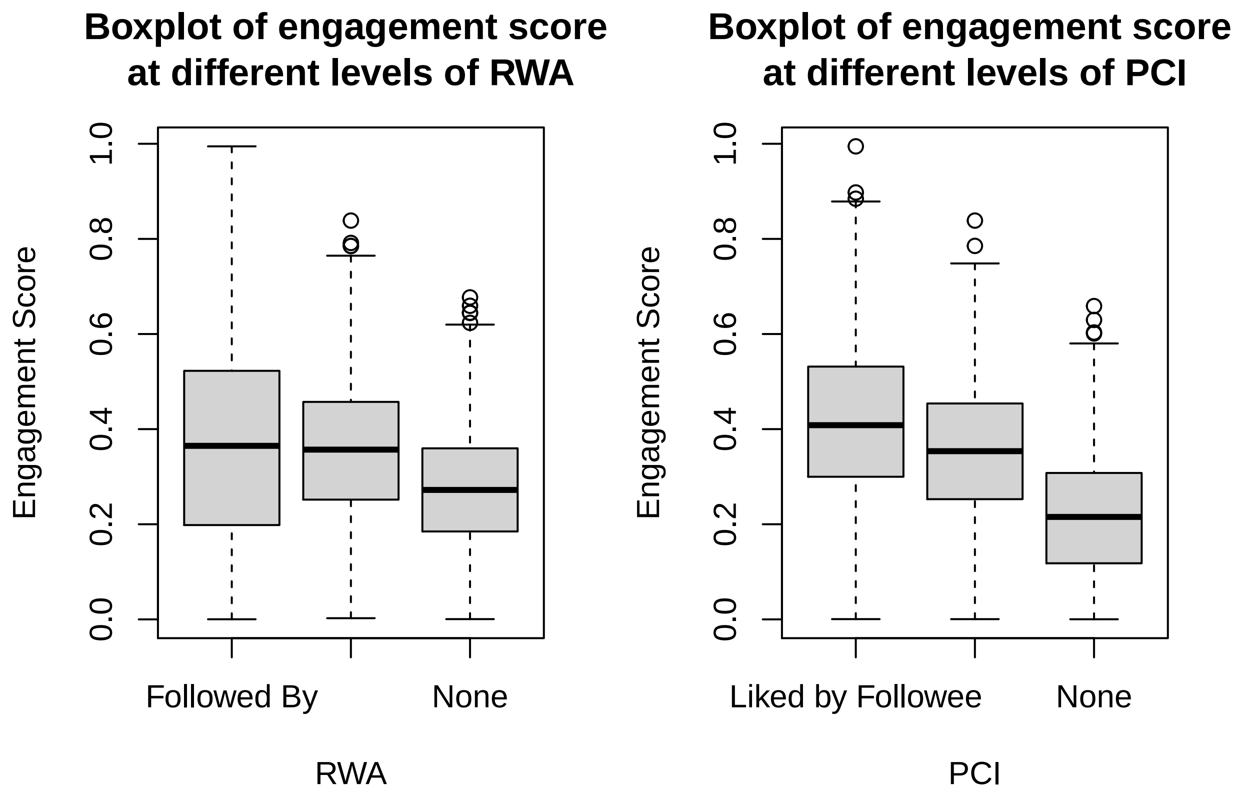
Test statistic: $t = 678.7588047$

p-value: $Pr(T \geq 678.7588) = 2.6368775 \times 10^{-258}$

Since the p-value $= 2.6368775 \times 10^{-258} < 0.05$, we reject the null hypothesis and conclude that the engagement score in each condition is not the same.

(f) Construct boxplots of the response variable vs. the levels of each nuisance factor. Does blocking appear to have been necessary? State YES or NO and provide a brief explanation.

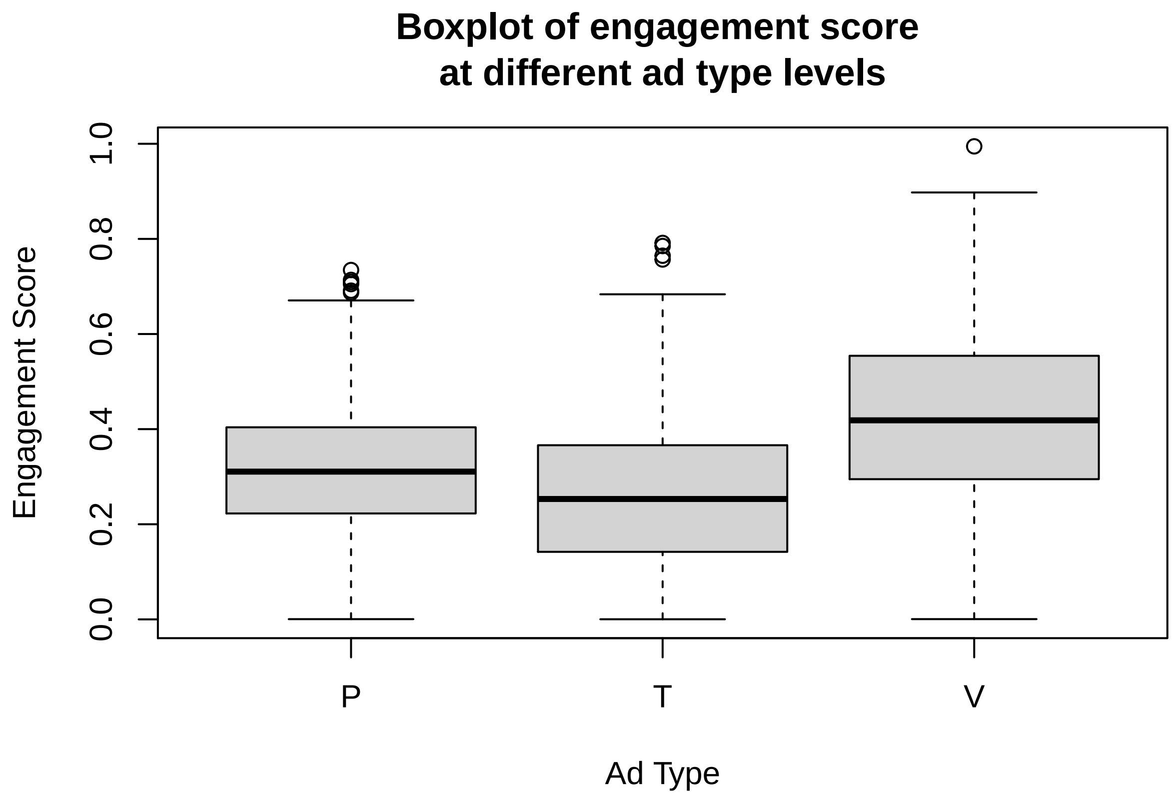
```
par(mfrow = c(1, 2))
boxplot(twitter$Engagement ~ twitter$RWA, xlab = "RWA", ylab = "Engagement Score")
title("Boxplot of engagement score \nat different levels of RWA")

boxplot(twitter$Engagement ~ twitter$PCI, xlab = "PCI", ylab = "Engagement Score")
title("Boxplot of engagement score \nat different levels of PCI")
```



Looking at the above boxplots, we can conclude that YES, blocking appears to have been necessary. This is reflected as we see that the engagement score appears to differ from one level to other for each of the nuisance factor.

(g) Construct a boxplot of the response variable vs. the levels of the design factor. Which level appears to be optimal?

```
boxplot(twitter$Engagement ~ twitter$Ad.Type, xlab = "Ad Type", ylab = "Engagement Score")
title("Boxplot of engagement score \nat different ad type levels")
```



From the above boxplot, we see that the engagement score is maximized when the type of ad shown is video. Hence video ad seems to be the optimal level.