Digital Advertising - A/B Test Report

Jing Li

Executive Summary

Game Fun is one of the world's top developers of casual mobile games and spends millions of dollars every year on digital advertising. In order to validate whether digital ads are effective, Game Fun ran an online display banners advertising campaign with the primary objective of increasing its sales on gaming subscription packages and conducted an A/B test to measure the causal effect.

Game Fun can run the promotion continuously according to the A/B test result and they should only focus on female gamers. Because only in this customer segment, the promotion ads effect on revenue/sales is positive (+16.4%).

However, the promotion ads did improve the purchase rate (4%) for all the customers. Game Fun could continue to conduct this campaign on all the customers only when their goal is to gain more customers regardless of the cost.

What's more, the promotion ads have raised 3.57% and 20.61% for the expected revenue of female gamers on site 1 and 2 respectively. It seems site 2 is more effective but we should take the ads cost into consideration. We can evaluate ROI for both sites to decide which one to focus on in the future.

Background

Game Fun is one of the world's top developers of casual mobile games and spends millions of dollars every year on digital advertising. Of particular interest to Game Fun is their efforts on improving their customer acquisition. In order to attract new users, Game Fun ran an online display banners advertising campaign with the primary objective of increasing its sales on gaming subscription packages.

Experiment Design

Before the start of the digital ad campaign, Game Fun chose two different websites ("content publishers") to run the experiment on. The content publishers have randomly assigned their web users to test and control groups. As users browsed on the two websites, the advertising servers checked whether a given user should be show a Game-Fun ad. If the user qualified for a Game-Fun ad, then the ad server checked whether the user was assigned to the test or the control group. If the user belonged to the test set, a Game-Fun ad was displayed to the user. Otherwise, a completely irrelevant ad was displayed to the user.

The ad advertised their most popular game and offered the user a promotion of $25 signing-up bonus. The credits would appear in the customer's game account and could be used to purchase any further in-app features. Based on historical data, a new customer subscription brings a revenue of $37.5 on average. This results in a net inflow of $12.5 after the $25 credit for the users acquired through this promotion.

However, Game-Fun had to pay the content publishers for these irrelevant ads, as well. This fact raised two concerns. First, paying for other companies' ads is directly decreasing their marketing budget. Second, they didn't like the fact that some users who saw an irrelevant ad might have signed up for the Game Fun game in the first place, had they been shown their gaming ad (indirect effect – opportunity cost missed).

For these two reasons, the data scientist team ran a statistical power analysis of the experiment and decided to allocate 70% of users to test group and the rest 30% to the control group, which controlled the opportunity cost and also maintained a statistically valid comparison.

Data Characteristics

The data analytics team collected experiment data as follows:

- Customer id
- Test (test group =1, control group = 0)
- Gender (male = 1, female = 0)
- Income (in thousands)
- Gamer (gamer = 1, non-gamer = 0)
- Site (two websites who publish the ads)
- Impressions (the number of advertising impressions that a customer received)
- Purchase (if the customer purchase anything within 30 days after conversion to the game; purchase = 1, no purchase = 0)

Experiment Results

1. ***Checking the Experiment Execution***

Before evaluating the effect of an experiment, it is important to make sure that the experiment was executed correctly.

test	0	1	diff_percentage
gamer	0.601823	0.601331	-0.081720
gender	0.647905	0.647289	-0.095049
income	55.166012	54.938236	-0.412890

Fig 1. Averages in test & control group for gamer, gender and income

By comparing the averages of income, gender and gamer variables in the test and control groups, it seems we have probabilistic equivalent samples. In order to make sure there is no statistically significant difference between test and control group, I conducted t-test (please see the result in jupyter notebook). The two-sample t-test outcome validated the hypothesis that there is no significant difference between average income/gamer/gender of the two groups. The result shows that in terms of these three aspects, different customers have equivalent chance to be selected in test or control group, which means we have probabilistic equivalent sample for the two groups. It guarantees the accuracy and efficiency of the A/B test.

2. ***Evaluating the Result***

- **Overall causal effect**

test	0	1	diff
purchase	0.036213	0.076822	0.040609

Fig 2. Purchase rate for all customers in test & control group

```
# For the purchasers in test group, the company has a cost of $25 for each one.
# So the revenue of each test purchaser is 12.5.
# The revenue of each control purchaser is 37.5 since they don't have the $25 credits.
# When calculating the relative increase percentage, I standardized the revenue according to the number of users
test_rev = gamefun[(gamefun['test']==1)]['purchase'].sum() * 12.5
control_rev = gamefun[(gamefun['test']==0)]['purchase'].sum() * 37.5
print("The expected total revenue of test group is",test_rev)
print("The expected total revenue of control group is",control_rev)
print("The relative percentage increase is", round((test_rev/7-control_rev/3)/(control_rev/3)*100,2) ,"%")
```

```
The expected total revenue of test group is 26975.0
The expected total revenue of control group is 16237.5
The relative percentage increase is -28.8 %
```

Fig 3. Expected Revenue for all customers in test & control group

The promotion ads increased purchase rate by around 4% and but decreased expected revenue by 28.8% for all the customers.

- **Customer segmentation**

1) Male vs. Female

test gender	0	1	diff
0	0.034442	0.080945	0.046503
1	0.037176	0.074575	0.037399

Fig 4. Purchase rate for different gender in test & control group

The promotion ads increased purchase rate by 4.7% for female and 3.7% for male.

2) Gamers vs. Non-gamers

test gamer	0	1	diff
0	0.037387	0.035092	-0.002295
1	0.035436	0.104487	0.069051

Fig 5. Purchase rate for gamers and non-gamers in test & control group

The promotion ads increased purchase rate by 7% for gamers but have a negative effect on non-gamers. So, we should focus on gamers to continue this promotion campaign.

3) Female Gamers vs Male Gamers

test	0	1	diff
gender			
0	0.032041	0.110092	0.078051
1	0.037275	0.101404	0.064129

Fig 6. Purchase rate for female gamers and male gamers in test & control group

The promotion ads increased purchase rate by 7.8% for female gamers and 6.4% for male gamers.

b. Comparison 4: Female Gamers vs Male Gamers

```
# Female Gamers
test_rev_fg = gamefun[(gamefun['test']==1) & (gamefun['gender']==0) & (gamefun['gamer']==1)]['purchase'].sum() * 12.5
control_rev_fg = gamefun[(gamefun['test']==0) & (gamefun['gender']==0) & (gamefun['gamer']==1)]['purchase'].sum() * 37.5
print("The expected total revenue of female gamers in test group is ",test_rev_fg)
print("The expected total revenue of female gamers in control group is ",control_rev_fg)
print("The relative percentage increase is", round((test_rev_fg/7-control_rev_fg/3)/(control_rev_fg/3)*100,2) ,"%")
```

```
The expected total revenue of female gamers in test group is  8250.0
The expected total revenue of female gamers in control group is  3037.5
The relative percentage increase is 16.4 %
```

```
# Male Gamers
test_rev_mg = gamefun[(gamefun['test']==1) & (gamefun['gender']==1) & (gamefun['gamer']==1)]['purchase'].sum() * 12.5
control_rev_mg = gamefun[(gamefun['test']==0) & (gamefun['gender']==1) & (gamefun['gamer']==1)]['purchase'].sum() * 37.5
print("The expected total revenue of male gamers in test group is ",test_rev_mg)
print("The expected total revenue of male gamers in control group is ",control_rev_mg)
print("The relative percentage increase is", round((test_rev_mg/7-control_rev_mg/3)/(control_rev_mg/3)*100,2) ,"%")
```

```
The expected total revenue of male gamers in test group is  13812.5
The expected total revenue of male gamers in control group is  6525.0
The relative percentage increase is -9.28 %
```

Fig 7. Expected revenue increase for female gamers and male gamers in test & control group

In terms of expected revenue, female gamers increased 16.4% but male gamers decreased by 9.28% which means the promotion ads have a negative effect on male gamers revenue.

4) Site 1 vs Site 2

test	0	1	diff
site			
site1	0.049261	0.146761	0.09750
site2	0.028746	0.102856	0.07411

Fig 8. Purchase rate for female gamers from the two sites in test & control group

The expected revenue

```
test_rev_1 = gamefun[(gamefun['test']==1) & (gamefun['site']=='site1') & (gamefun['gamer']==1) & (gamefun['gender']==0)]['pu
control_rev_1 = gamefun[(gamefun['test']==0) & (gamefun['site']=='site1') & (gamefun['gamer']==1) & (gamefun['gender']==0)][
print("The expected revenue of site 1 female gamers in test group is ",test_rev_1)
print("The expected revenue of site 1 female gamers in control group is ",control_rev_1)
print("The relative percentage increase is", round((test_rev_1/7-control_rev_1/3)/(control_rev_1/3)*100,2) ,"%")
```

```
The expected revenue of site 1 female gamers in test group is  1812.5
The expected revenue of site 1 female gamers in control group is  750.0
The relative percentage increase is 3.57 %
```

```
_rev_2 = gamefun[(gamefun['test']==1) & (gamefun['site']=='site2') & (gamefun['gamer']==1) & (gamefun['gender']==0)]['purcha
rol_rev_2 = gamefun[(gamefun['test']==0) & (gamefun['site']=='site2') & (gamefun['gamer']==1) & (gamefun['gender']==0)]['pur
t("The expected revenue of site 2 female gamers in test group is ",test_rev_2)
t("The expected revenue of site 2 female gamers in control group is ",control_rev_2)
t("The relative percentage increase is", round((test_rev_2/7-control_rev_2/3)/(control_rev_2/3)*100,2) ,"%")
```

```
The expected revenue of site 2 female gamers in test group is  6437.5
The expected revenue of site 2 female gamers in control group is  2287.5
The relative percentage increase is 20.61 %
```

Fig 9. Expected revenue for female gamers from the two sites in test & control group

As we can see, the promotion effect on the purchase rate for female gamers (as we have decided to focus on female gamers) in the two sites is 9.7% and 7.4% respectively. In addition, the promotion has raised 3.57% and 20.61% for the expected revenue on site 1 and 2 respectively.

<u>Conclusion & Recommendation</u>

In conclusion, Game Fun should run the promotion in the future, but they should only focus on female gamers.

According to the A/B test result, the purchase rate increased 4% after promotion for all customers. However, in terms of the expected revenue, it decreased around 28.8% which indicates the promotion ads is not profitable.

From the analysis result of part 2, the promotion effect on purchase rate of male vs. female is similar. But for gamers vs. non-gamers, the test and control difference for non-gamers is negative which means the promotion actually discouraged them to purchase. It's understandable that people don't play games feels annoyed when we show too much game ads to them. So, we should only focus on gamers to conduct the promotion ads.

From a deeper analysis, it turns out the promotion effect on purchase rate for the male gamers and female gamers are similar (6.4% vs. 7.8%). However, it results in 16.4% increase for female gamers but 9.28% decrease for male gamers in terms of expected revenue. Consequently, Game-Fun should only focus on female gamers for promotion ads in the future.

Additionally, as we can see the promotion ads has raised 3.57% and 20.61% for the expected revenue on site 1 and 2 respectively. We could say that the promotion ads on site 2 is more effective. However, it’s important to take the ads cost into consideration. We could evaluate ROI from both sites to make the final decision.

Limitation

One of the limitations about this experiment is that sometimes we run t-test before executing an experiment and find a large difference between test and control groups which means we can't use randomization to create probabilistic equivalent groups. In that case, we can try to create "functionally equivalent" groups through matching. We only consider equivalent on the observable characteristics like income, gender and gamer. We do not care about the unobservable ones. The assumption is when the objects are similar on observable characteristics, they are also similar on unobservable characteristics. We match objects with similar observable characteristics goes to A and B group respectively, drop those without reasonable match. Doing exact matching is impractical, we can try propensity score matching. Another limitation is that we cannot apply significance test on big data as the sample size is too large, we will always get a significant result in the end. One possibility for addressing this problem is to let the critical value be a function of the sample size. We could think about this as a testing problem for nested models. Cameron and Trivedi (2005) suggest using the Bayesian Information Criterion (BIC) for which the penalty increases with the sample size. Using the BIC for testing the significance of one variable is identical to using a two-sided t-test critical value of sqrt(ln(N)).