Globox A/B Hypothesis Test Written report

Tableau visualizations- Link
G sheets with data -Link
G Slides Presentation- Link
Video Presentation -Link

Understanding the data

In order to complete a thorough and sound hypothesis test for the users who saw the banner and those who didn't we first need to understand the data. We decided to design a quick ERD to describe the schema used in the database. We then spent a good amount of time looking at the data in the database before extracting the necessary data in a concise query. We decided to only extract the sum of the amount spent per user in each group as well as the dates. This data was downloaded as a csv and brought in to google sheets.

```
SELECT g.uid,
       g.group,
       join_dt,
       COALESCE(Sum(spent), 0) AS total_spent
FROM    activity AS a
       FULL JOIN groups AS g
              ON g.uid = a.uid
GROUP   BY 1,
          2,
          3
ORDER   BY 1
```

Preparation for analysis

After first formatting the data column and using named ranges to make the aggregations quicker, we made a pivot table to quickly find the measures of central tendency like minimum / maximum value, the average, and the total amount in each group. We discovered there were fewer people in the control and on average people spent $.016 more in the treatment while the standard deviation was almost identical.

This gives us a quick understanding of the subtle differences in the values of each group but it's not enough to base our decisions on since this is only a small sample of our total user base.

Hypothesis test for sample means

We go on to set up our hypothesis test. We first state our null hypothesis; (the assumed statement that we seek to disprove) that there is no difference in the average amount spent per user in the population (total user base) between the users who saw the banner and those who didn't. We then make our alternate hypothesis (a statement that could explain the data should the null hypothesis be rejected), which is that there is a difference in the average amount spent per user in the treatment, possibly due to seeing the banner. We set a significance level of 5% to say that we are 95% sure that our data results are a true representation of the untested users in the population. We separate the users by group using a pivot table in Sheets and add the values as the amount spent. We use the ttest function in sheets to calculate our p-value and find a p-value of .944 which is obviously more than our significance level and so we fail to reject our null hypothesis. This doesn't prove the null hypothesis. It just means we are 95% there is no difference between the amount spent between the control and the treatment groups.

Hypothesis test for conversion rate

We go on to do a similar hypothesis test for the conversion rate. First we find the amount of people who spent money in each group and divide that by the total amount of people in each group. This gives us the conversion rate. Now we see there is a .71% increase in the treatment group but we don't know if this is due to sampling error or if there is an actual difference in the population.

We perform a hypothesis test for a difference in proportions. We set the same significance level of 5% and a confidence level of 95% and we begin by pooling our proportions together and finding our standard error, z score and our p-value. We discover that the p-value is almost 0! Which tells us that there is almost no chance that we would find the conversion rates that we calculated if the null hypothesis was true. And since we set our significance level to .05 and our p-value is much smaller than that we can reject our null hypothesis.

Confidence Intervals

Next we calculate the confidence interval for our sample means and conversion rate which tells us the range where our true population means and conversion rate can fall in relation to our sample. We plot these in a couple charts in our google sheets document. The intervals overlap in the sample mean, but do not overlap in the conversion rate. This document is downloaded as a csv.

Tableau visuals

The csv file is brought into tableau and used to make 2 simple line graphs depicting the trend of metrics over the 12 days of the test. The first metric is the total amount spent and the second is the number of conversions. These are filtered by the control and the treatment group to see the differences between the trends. We see from these graphs that there is a strong negative trend over time, particularly in the first few days of the test. A forecast is then made for the next few days after the test and this shows a continuation of the negative trend.
These visualizations are put into a dashboard for ease of presentation and a live filter is placed for the group so a user can choose which group's trend they want to see visualized.

Conclusion & Recommendations

We first need to understand what was concluded by these tests and what wasn't. We determined to see with 95% confidence whether the statistics we found in our sample were representative of the larger user population. We saw a difference in both the average amount spent and the conversion rate. But from our hypothesis test we could only find that the conversion rate was likely a true representation of the population. We cannot say the difference in the sample mean was due to a fluke or any extenuating circumstances. We also must admit there is still a possibility that the findings in the conversion rate could also be inaccurate even with a 95% accuracy. This is called a Type I error.

There are many things we can recommend going forward. Firstly it's important to know why both metrics dropped off so steeply after starting the test. Especially until about Jan 28-29th this could inform our future decisions. We can also recommend further analysis into the churn rate of both groups and whether they are returning users. Some more information into the demographics of the users, where they come from and age groups could prove useful. Also a similar test for the desktop users might prove useful as this was only tested on the mobile application. And since we saw almost no difference in the amount spent between groups perhaps an heightened awareness of hot sale items or popular items into the design of the banner might jumpstart sales. So as it stands we would recommend a no launch until some of these considerations are met, though we are not too far off from a launch.