A/B Testing Analysis using Python

A/B testing is a statistical technique - a randomized controlled experiment where two or more versions of a variable (such as a landing page, page element, marketing campaign, etc.) are shown to different groups of users and compared to see which one performs better and drives the success metric. There are controlled and treatment groups with a sufficiently large sample size to ensure that the difference in the metric (if any) is statistically significant, with no possible confounding factors.

Scenario

At a medium-sized online e-commerce business, the UX designer worked on a new version of the product page, with the objective of increasing conversion rate. The current conversion rate is about 13% on average throughout the year, and that the team is targeting a lift of 2%, meaning that the new design will be considered a success if it raises the conversion rate to 15%. Before rolling out the change, the team conducts A/B test on a subset of users to see how well the new page is performing.

Dataset

- The dataset has been obtained from Kaggle; it includes the results of an A/B test performed for two different versions of a webpage.
- It includes details about the customer *(user_id)*, time *(timestamp)*, whether the user belonged to a control or treatment group *(group)*, if they saw the old or new webpage *(landing page)* and finally if they made a purchase or not *(converted)*.

user_id	timestamp	group	landing_page	converted
851104	11:48.6	control	old_page	0
804228	01:45.2	control	old_page	0
661590	55:06.2	treatment	new_page	0
853541	28:03.1	treatment	new_page	0
864975	52:26.2	control	old_page	1
936923	20:49.1	control	old_page	0
679687	26:46.9	treatment	new_page	1
719014	48:29.5	control	old_page	0
817355	58:09.0	treatment	new_page	1
839785	11:06.6	treatment	new_page	1
929503	37:11.5	treatment	new_page	0
834487	37:47.8	treatment	new_page	0
803683	05:16.2	treatment	new_page	0

Steps for analysis

1. **Experiment Design**

   a. Hypothesis formulation

   It is a two tailed test as we don't know if the new design will perform better or worse than the old one. So the null hypothesis is that the new design is performing the same old design. The confidence level has been set at 95%. p and $\boldsymbol{p_o}$ stand for the conversion rate of the new and old design, respectively.

   H_o: $p = p_o$
   H_a: $p \neq p_o$
   $\alpha = 0.05$

   b. Selecting the variables

   - Independent variable: Group (Control v/s Treatment)
   - Dependent variable: Conversion rate

   Having both control and treatment groups will help us control affect of other variables on the final result. **The control group will see the old page while the treatment group will see the new page.** Conversion rate will be measured using the binary converted attribute. 0's stand for users who didn't make a purchase during the session and 1's are for users who made a purchase.

   c. Calculating sample size

   To ensure that the sample represents our user population and is large enough to detect a statistically significant difference in the two groups, power analysis has been used to calculate the sample size.

   Power parameter has been set as 0.8, alpha as 0.05, and the effect size is between 13% and 15%.

2. **Data Exploration and preparation**

   Unlike in the real world, where the data will be collected after the above step is complete, here we already have the dataset from Kaggle. The number of rows were reduced to match our sample size.

   In this step we used jupyter notebook to load the data, perform exploratory steps like calculating number of rows, columns, checking the data type, and finding & removing duplicate values. Followed by exploration and cleaning, the data was then sampled to have 4720 entries from each group.

3. **Analysis & Visualization**

In this step, we calculated the conversion rates, standard deviation and error for both control and treatment groups and found that the conversion rates of both the groups were very close.

4. **Testing the hypothesis**

We used z-test for calculating our p-value and got the below results:

- z statistic: -0.34
- p-value: 0.732
- CI 95% for control group: [0.114, 0.133]
- CI 95% for treatment group: [0.116, 0.135]

Conclusion

The p-value is 0.732. Since, this is higher than the alpha = 0.05, we cannot reject the null hypothesis. This means that there is no significant difference between the conversion rates resulted from old v/s new landing pages.

We can conclude that the new page didn't perform any better than the old page.