

Assignment 4

Understand The Concept of A/B Testing and
How To Analyze Its Result

By Sarah Fatria Mumtaz
DS03286

Assessment Requirement

- Problem definition
 - Hypothesis formulation
 - Hypothesis testing
 - Result interpretation
-

Problem Definition

An e-commerce company has developed a new web page in order to try and increase the number of users who "convert", meaning the number of users who decide to pay for the company's product. So, the goal here is to determine whether the version of the website affects the customer's tendency to buy company's products or not by implementing a/b testing.

Hypothesis Formulation

While implementing chi square, we have to create two hypotheses i.e Null Hypothesis and Alternate Hypothesis. Since the problem is *does the version of the website affect customer's tendency to buy company's products or not*, the hypotheses could be like this:

- Null Hypothesis: there is no relation between page version and customer's tendency to purchase products. they are independent
- Alternate Hypothesis: there is a relation between page version and customer's tendency to purchase products. they are dependent

Hypothesis Testing

In this step, I read the data to df and display it. It shows that df has five columns, which is user_id, timestamp, group, landing_page, and converted. From these column, I only use the landing_page and converted to solve the problem

This step displays the unique values of landing_page, that is old_page and new_page

Then, I set the acceptance value is 0.05 and display the count of user who convert and not from landing_page in observed_values variables. From this step, we can count the old page and new page impact rate using this formula:

Version_impact_rate =
$$(\text{total_user_who_convert}) / (\text{total user})$$

```
#read the data and display it
df = pd.read_csv('/content/drive/MyDrive/ecommerce_ab_testing/ab_data.csv')
df
```

	user_id	timestamp	group	landing_page	converted
0	851104	11:48.6	control	old_page	0
1	804228	01:45.2	control	old_page	0

```
df.landing_page.unique()
```

```
array(['old_page', 'new_page'], dtype=object)
```

```
#if the p-value is < 0.05 then we reject the null_hypo and accept the alter_hypo.
#else if the p-value is >= 0.05 then we'll do the opposite
acceptance_criteria = 0.05
```

```
#we'll count the rate. first, we will see the sum and count using pd.crosstab
observed_values = pd.crosstab(df["landing_page"], df["converted"]).values
observed_values
```

```
array([[129743, 17498],
       [129500, 17739]])
```

Hypothesis Testing

```
#implementing the chi square
chi2_statistic, p_value, dof, expected_values = chi2_contingency(observed_values, correction = False)
print("dof value: ", dof)
print("chi2_statistic value: ", chi2_statistic)
print("p_value: ", p_value)

#find the critical value
#The critical value = for any per-determined probability (p), the test would indicate a result that is less probable than p.
critical_value = chi2.ppf(1 - acceptance_value, dof)
print("critical value: ", critical_value)

#implementing the ab testing using acceptance_value
if p_value < 0.05:
    print("since the p_value is ", p_value, " which is < 0.05, the null hypothesis should be rejected and alternate hypothesis should be accepted")
    print("we'll accept this hypothesis: " + (alter_hypo))
else:
    print("since the p_value is ", p_value, " which is >= 0.05, the null hypothesis should be accepted and alternate hypothesis should be rejected")
    print("we'll accept this hypothesis: " + null_hypo)
```

dof value: 1
chi2_statistic value: 1.87605695553178
p_value: 0.17078297802593548
critical value: 3.841458820694124
since the p_value is 0.17078297802593548 which is >= 0.05, the null hypothesis should be accepted and alternate hypothesis should be rejected
we'll accept this hypothesis: there is no relation between page version and customer's decision to purchase products. they are independent

Then, I implement the a/b testing. If the p_value is less than the acceptance_value, then the Null Hypothesis will be rejected and Alternate Hypothesis will be accepted. But, if the p_value is not less than acceptance_value, the opposite will happen.

Result

From the experiment, I get a result that the p value is 0.17078297802593548, which is not less than 0.05. Therefore, the Null Hypothesis will be accepted and Alternate Hypothesis will be rejected. The accepted hypothesis is “There is no relation between page version and customer's tendency to purchase products. They are independent”. In other words, the company should create different kind of webpage or maybe make other changes to increase the customer's tendency to buy their product.

The dataset:

<https://www.kaggle.com/datasets/putdejdomthai/ecommerce-ab-testing-2022-dataset1>

See the full version code here:

<https://colab.research.google.com/drive/1RaIpHuP2JL7Hr61hTTT95ILJ4uQjQB2W?usp=sharing>