

# A/B Test with Python

Yasemin Turker

2022



# INTRODUCTION

---

## Problem

An advertising company has developed a new ad to have users engage with their questionnaire.

The company has shown the new ad to some users and a dummy ad to others and wants their data analyst team to interpret the results.

**Does the new ad generate more responses to their questionnaire?**

**Is it statistically significant?**



**Is the company justified in using the new ad?**



“



*A/B testing is a type of split testing and is commonly used to drive improvements to specific variables or elements by measuring user or audience engagement.<sup>1</sup>*

# DATA

---

Dataset is found on kaggle from an advertising company:

<https://www.kaggle.com/osuolaleemmanuel/ad-ab-testing>

$$H_0: \mu_{\text{con}} = \mu_{\text{exp}}$$

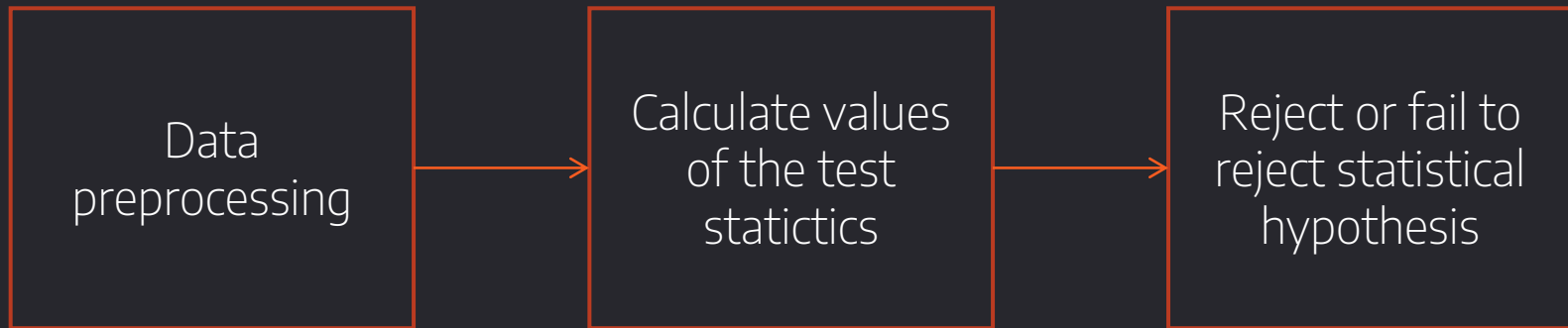
Null Hypothesis : There is no significant difference between the ad success rate of both groups

$$H_1: \mu_{\text{con}} \neq \mu_{\text{exp}}$$

Alternative Hypothesis : There is significant difference between the ad success rate of both groups

# A/B TESTING PROCESS

---



# DATA PREPROCESSING

- Check for null values
- Remove non-answer observations





# DATA PREPROCESSING

## Checked for null values

```
# Checking for nulls  
Total_nulls = ABdata.isnull().sum()  
print('Null values:')  
print(Total_nulls)
```

```
Null values:  
auction_id      0  
experiment      0  
date            0  
hour            0  
device_make     0  
platform_os     0  
browser         0  
yes             0  
no              0  
dtype: int64
```

# DATA PREPROCESSING

## Removed non-answer observations

```
# non-answer observations (both 'yes' and 'no' columns are equal to 0)
ABdata_00 = ABdata[(ABdata['yes'] == 0) & (ABdata['no'] == 0)]
ABdata_00
```

	auction_id	experiment	date	hour	device_make	platform_os	browser	yes	no
0	0008ef63-77a7-448b-bd1e-075f42c55e39	exposed	2020-07-10	8	Generic Smartphone	6	Chrome Mobile	0	0
1	000eabc5-17ce-4137-8efe-44734d914446	exposed	2020-07-07	10	Generic Smartphone	6	Chrome Mobile	0	0
3	00187412-2932-4542-a8ef-3633901c98d9	control	2020-07-03	15	Samsung SM-A705FN	6	Facebook	0	0
4	001a7785-d3fe-4e11-a344-c8735acacc2c	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	0
5	0027ce48-d3c6-4935-bb12-dfb5d5627857	control	2020-07-03	15	Samsung SM-G960F	6	Facebook	0	0
...	...	...	...	...	...	...	...	...	...
8072	ffea24ec-cec1-43fb-b1d1-8f93828c2be2	exposed	2020-07-05	7	Generic Smartphone	6	Chrome Mobile	0	0
8073	ffea3210-2c3e-426f-a77d-0aa72e73b20f	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	0
8074	ffea0f1-1d72-4ba9-afb4-314b3b00a7c7	control	2020-07-04	9	Generic Smartphone	6	Chrome Mobile	0	0
8075	ffeed62-3f7c-4a6e-8ba7-95d303d40969	exposed	2020-07-05	15	Samsung SM-A515F	6	Samsung Internet	0	0
8076	ffbb9ff-568a-41a5-a0c3-6866592f80d8	control	2020-07-10	14	Samsung SM-G960F	6	Facebook	0	0

6834 rows × 9 columns

# DATA PREPROCESSING

## Removed non-answer observations

```
# dropped non-answer observations
ABdata_A00= ABdata.drop(ABdata_00.index)
print('# Observations: {}'.format(ABdata_A00.shape[0]))
ABdata_A00.head()
```

```
# Observations: 1243
```

	auction_id	experiment	date	hour	device_make	platform_os	browser	yes	no
2	0016d14a-ae18-4a02-a204-6ba53b52f2ed	exposed	2020-07-05	2	E5823	6	Chrome Mobile WebView	0	1
16	008aafdf-deef-4482-8fec-d98e3da054da	exposed	2020-07-04	16	Generic Smartphone	6	Chrome Mobile	1	0
20	00a1384a-5118-4d1b-925b-6cdada50318d	exposed	2020-07-06	8	Generic Smartphone	6	Chrome Mobile	0	1
23	00b6fadb-10bd-49e3-a778-290da82f7a8d	control	2020-07-08	4	Samsung SM-A202F	6	Facebook	1	0
27	00ebf4a8-060f-4b99-93ac-c62724399483	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	1

# EXPLORATORY DATA ANALYSIS

- Process of performing some initial investigations on the dataset



# EXPLORATORY DATA ANALYSIS

```
NumberOfControl=ABdata_A00['experiment'].value_counts().control  
NumberOfExposed=ABdata_A00['experiment'].value_counts().exposed  
NumberOfControl,NumberOfExposed
```

(586, 657)

```
Control_yes= ExperimentsYes.iloc[0]  
Exposed_yes= ExperimentsYes.iloc[1]  
Control_yes,Exposed_yes
```

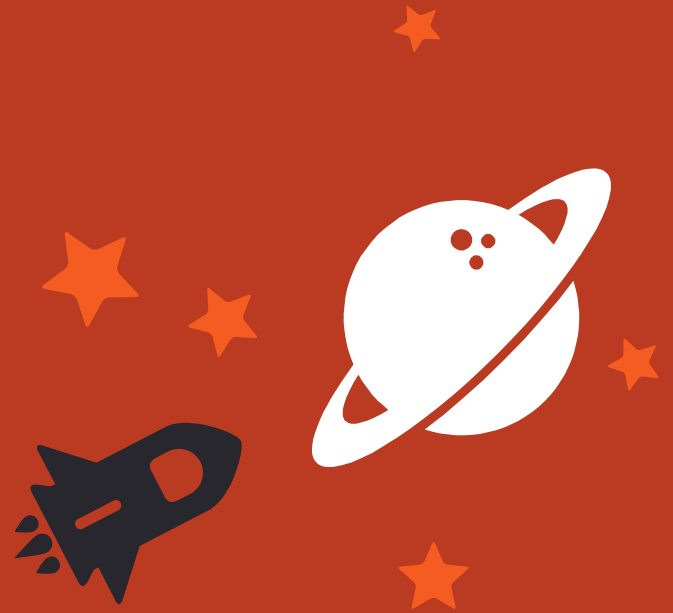
(264, 308)

```
RateOfControl= Control_yes/NumberOfControl  
RateOfExposed= Exposed_yes/NumberOfExposed  
RateOfControl, RateOfExposed
```

(0.45051194539249145, 0.4687975646879756)

# 2 SAMPLE Z TEST

- Investigate the significant differences between the control and experimental groups' metrics



# 2 SAMPLE Z-TEST

Test Stat.	Z Critical	P Value	Confidence Interval
-0.6457	1.9599	1.4815	[-0.07380 0.03722]

## Conclusion

Test stat:0.6457 < Z Crit:1.96  
can not be rejected the null hypothesis

# CONCLUSIONS

---

- The null hypothesis can not be rejected and conclude that there is no statistically significant difference between the AdA and AdB.
- The company is not justified in using the new ad because the findings show that the new ad will not provide any additional benefit.



# THANKS!

ANY QUESTIONS?

- [yaseminaturker@gmail.com](mailto:yaseminaturker@gmail.com)



# REFERENCES

---

1- <https://www.seldon.io/a-b-testing-for-machine-learning>